HEAVY TAIL PHENOMENA IN PREFERENTIAL ATTACHMENT NETWORKS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Tiandong Wang

August 2019

HEAVY TAIL PHENOMENA IN PREFERENTIAL ATTACHMENT NETWORKS

Tiandong Wang, Ph.D.

Cornell University 2019

Preferential attachment is widely used to model the power-law behavior of degree distributions in social networks. In this thesis, we study three aspects of a directed preferential attachment model. First, we consider fitting this network model under different data scenarios. We propose both parametric and semi-parametric estimation procedures and compare the corresponding estimating results. Second, we see from empirical studies that statistical estimates of the marginal tail exponent of the power-law degree distribution often use the Hill estimator, even though no theoretical justification has been given. Hence, we study the convergence of the joint empirical measure for in- and out-degrees and prove the consistency of the Hill estimator for the preferential attachment model. Finally, we consider a widely adopted threshold selection procedure when estimating the power-law index in practice and examine the asymptotic behavior of the selected threshold as well as the corresponding power-law index given.

BIOGRAPHICAL SKETCH

Tiandong Wang was born in Maanshan, an industrial city located in Eastern China stretching across the Yangtze River. After finishing high school, she went to the Australian National University in Canberra, Australia, where she obtained her bachelor's degree in actuarial science with honors in statistics. After five years in Australia, she decided to go to the U.S. and attended Cornell for Ph.D. study.

During the five years in Ithaca, Tiandong appreciates the beautiful scenery and peaceful environment in this small town. She treasures the life there and enjoys travels to other American cities during spring and fall breaks. Upon graduating from Cornell, Tiandong will join the Department of Statistics at Texas A&M University as an assistant professor.

To my grandma, Yongsu Chen.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Prof. Sidney Resnick, for his patient guidance, constant support and enlightening advice throughout my Ph.D. studies. This thesis would not have been possible without his direction. I have learned from him not just his immense knowledge in research but also his humor and enthusiasm towards life. I have been extremely lucky to have an advisor who cares so much about my work and is always responsible.

I would like to extend my sincere gratitude to Prof. Richard Davis and Prof. Gennady Samorodnitsky for their mentorship as well as strong support during my job search. I thank Dr. Phyllis Wan, Prof. Holger Drees and Prof. Anja Janßen for being great collaborators. I am also grateful to Prof. Robert Jarrow and Prof. Martin Wells for being my special committee members.

Completing the Ph.D. study would have been much harder without the support and friendship provided by my Ph.D. cohort. A big thank you goes to my friends at Cornell, especially Xiaoyun Quan, Sijia Ma, Jiayang Gao, Yixuan Zhao, Yuhang Ma and Pu Yang, who have always been a source of help for my life as a graduate student. I thank Jingyu Huang, Chenyao Hu, Dan Wang and Shuyu Zhu for our life-long friendship.

I thank the School of Operations Research and Information Engineering for providing me with such a precious opportunity to undertake this research and get to know so many brilliant people.

Last but not the least, I owe my deepest gratitude to my parents, Yuzhang Wang and Wanning Du, who always give me unconditional support and encouragement. This thesis is also dedicated to the memory of my grandmother, Yongsu Chen. You are gone but your belief in me has made this journey possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

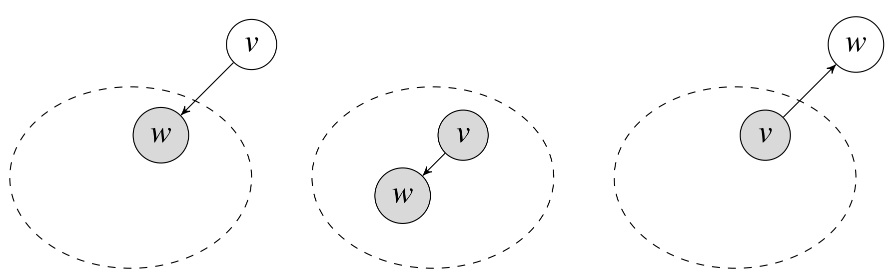
1.1 Background

The preferential attachment (PA) mechanism, in which edges and nodes are added to the network based on probabilistic rules, provides an appealing description for the evolution of a network. The rule for how edges connect nodes depends on node degree; large degree nodes attract more edges. The idea is applicable to both directed and undirected graphs and is often the basis for studying social networks, collaborator and citation networks, and recommender networks. Elementary descriptions of the preferential attachment model can be found in [20] while more mathematical treatments are available in [5, 19, 59]. Also see [35] for a statistical survey of methods for network data, [53] for consideration of statistics of an undirected network and [71] for asymptotics of a directed exponential random graph models. Limit theory for estimates of an undirected preferential attachment model was considered in [23].

For many networks, empirical evidence supports the hypothesis that in- and out-degree distributions follow a power law. This property was shown to hold in linear preferential attachment models, which made preferential attachment an attractive choice for network modeling [7, 19, 39, 40, 59]. While the marginal degree power laws in a simple linear preferential attachment model were studied in [7, 39, 40], the joint regular variation (see [50, 51]) which was akin to a *joint power law*, was only recently established [52, 55].

1.1.1 Model Specification

The directed edge preferential attachment model [7, 40] constructs a growing directed random graph $G(n) = (V(n), E(n))$ whose dynamics depend on five non-negative real numbers α, β, γ, δ_{in} and δ_{out}, where $\alpha + \beta + \gamma = 1$ and $\delta_{\text{in}}, \delta_{\text{out}} > 0$. To avoid degenerate situations, assume that each of the numbers α, β, γ is strictly smaller than 1. We obtain a new graph $G(n)$ by adding one edge to the existing graph $G(n-1)$ and index the constructed graphs by the number n of edges in $E(n)$. We start with an arbitrary initial finite directed graph $G(n_0)$ with at least one node and n_0 edges. For $n > n_0$, $G(n) = (V(n), E(n))$ is a graph with $|E(n)| = n$ edges and a random number $|V(n)| = N(n)$ of nodes. If $u \in V(n)$, $D_{\text{in}}^{(n)}(u)$ and $D_{\text{out}}^{(n)}(u)$ denote the in- and out-degree of u respectively in $G(n)$. There are three scenarios that we call the α, β and γ-schemes, which are activated by flipping a 3-sided coin whose outcomes are $1, 2, 3$ with probabilities α, β, γ. More formally, we have an iid sequence of multinomial random variables $\{J_n, n > n_0\}$ with cells labelled $1, 2, 3$ and cell probabilities α, β, γ. Then the graph $G(n)$ is obtained from $G(n-1)$ as follows.



α-scheme $\qquad$ β-scheme $\qquad$ γ-scheme

- If $J_n = 1$ (with probability α), append to $G(n-1)$ a new node $v \in V(n) \setminus V(n-1)$ and an edge (v, w) leading from v to an existing node $w \in V(n-1)$. Choose the existing node $w \in V(n-1)$ with probability depending on its in-degree in $G(n-1)$:

$$\mathbf{P}[\text{choose } w \in V(n-1)] = \frac{D_{\text{in}}^{(n-1)}(w) + \delta_{\text{in}}}{n-1+\delta_{\text{in}} N(n-1)}. \tag{1.1.1}$$

- If $J_n = 2$ (with probability β), add a directed edge (v, w) to $E(n-1)$ with $v \in V(n-1) = V(n)$ and $w \in V(n-1) = V(n)$ and the existing nodes v, w are chosen independently from the nodes of $G(n-1)$ with probabilities

$$\mathbf{P}[\text{choose } (v,w)] = \Big(\frac{D_{\text{out}}^{(n-1)}(v) + \delta_{out}}{n-1+\delta_{out} N(n-1)}\Big)\Big(\frac{D_{\text{in}}^{(n-1)}(w) + \delta_{in}}{n-1+\delta_{\text{in}} N(n-1)}\Big).$$

- If $J_n = 3$ (with probability γ), append to $G(n-1)$ a new node $w \in V(n) \setminus V(n-1)$ and an edge (v, w) leading from the existing node $v \in V(n-1)$ to the new node w. Choose the existing node $v \in V(n-1)$ with probability

$$\mathbf{P}[\text{choose } v \in V(n-1)] = \frac{D_{\text{out}}^{(n-1)}(v) + \delta_{\text{out}}}{n-1+\delta_{out} N(n-1)}. \tag{1.1.2}$$

Note that this construction allows the possibility of having self loops in the case where $J_n = 2$, but the proportion of edges that are self loops goes to 0 as $n \to \infty$. Also, multiple edges are allowed between two nodes.

1.1.2 Power law of degree distributions

Given an observed network with n edges, let $N_{ij}(n)$ denote the number of nodes with in-degree i and out-degree j. If the network is generated from the linear preferential attachment model described above, then from [7], there exists a proper probability distribution $\{f_{ij}\}$ such that almost surely

$$\frac{N_{ij}(n)}{N(n)} \to f_{ij} =: \frac{p_{ij}}{1-\beta}, \quad n \to \infty. \tag{1.1.3}$$

Consider the limiting marginal in-degree distribution $p_i^{\text{in}} := \sum_j p_{ij}$. The authors show in [7, Equation (3.10)] that

$$p_0^{\text{in}} = \frac{\alpha}{1 + a_1(\delta_{\text{in}})\delta_{\text{in}}},$$

and for $i \geq 1$,

$$p_i^{\text{in}} = \frac{\Gamma(i + \delta_{\text{in}})\Gamma(1 + \delta_{\text{in}} + a_1(\delta_{\text{in}})^{-1})}{\Gamma(i + 1 + \delta_{\text{in}} + a_1(\delta_{\text{in}})^{-1})\Gamma(1 + \delta_{\text{in}})} \left(\frac{\alpha\delta_{\text{in}}}{1 + a_1(\delta_{\text{in}})\delta_{\text{in}}} + \frac{\gamma}{a_1(\delta_{\text{in}})} \right),$$

where

$$a_1(\lambda) := \frac{\alpha + \beta}{1 + \lambda(1 - \beta)}, \quad \lambda > 0.$$

Moreover, p_i^{in} satisfies

$$p_i^{\text{in}} := \sum_{j=0}^{\infty} p_{ij} \sim C_{\text{in}} i^{-(1+\iota_{\text{in}})} \text{ as } i \to \infty, \quad \text{as long as } \alpha\delta_{\text{in}} + \gamma > 0, \tag{1.1.4}$$

for some finite positive constant C_{in}, where the power index

$$\iota_{\text{in}} = \frac{1 + \delta_{\text{in}}(\alpha + \gamma)}{\alpha + \beta} \tag{1.1.5}$$

Similarly, the limiting marginal out-degree distribution has the same property:

$$p_j^{\text{out}} := \sum_{i=0}^{\infty} p_{ij} \sim C_{\text{out}} i^{-(1+\iota_{\text{out}})} \text{ as } j \to \infty, \quad \text{as long as } \gamma\delta_{\text{out}} + \alpha > 0,$$

for C_{out} positive and

$$\iota_{\text{out}} = \frac{1 + \delta_{\text{out}}(\alpha + \gamma)}{\beta + \gamma}. \tag{1.1.6}$$

Let (I, O) be a fictitious random vector with joint pmf p_{ij}, then

$$\mathbf{P}(I \geq i) \sim C_{\text{in}}(1 + \iota_{\text{in}})^{-1} \cdot i^{-\iota_{\text{in}}} \quad \text{as } i \to \infty, \tag{1.1.7}$$

$$\mathbf{P}(O \geq j) \sim C_{\text{out}}(1 + \iota_{\text{out}})^{-1} \cdot j^{-\iota_{\text{out}}} \quad \text{as } j \to \infty. \tag{1.1.8}$$

In the linear PA model, the joint distribution of (I, O) satisfies non-standard regular variation. Let $\mathbb{M}(\mathbb{R}_+^2 \setminus \{\mathbf{0}\})$ be the set of Borel measures on $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$ that are finite on sets bounded away from the origin. Then (I, O) is *non-standard regularly varying* on $\mathbb{R}_+^2 \setminus \{\mathbf{0}\}$ means that as $t \to \infty$,

$$t\mathbf{P}\left[\left(\frac{I}{t^{1/\iota_{\text{in}}}}, \frac{O}{t^{1/\iota_{\text{out}}}}\right) \in \cdot\right] \to \nu(\cdot), \quad \text{in } \mathbb{M}(\mathbb{R}_+^2 \setminus \{\mathbf{0}\}), \tag{1.1.9}$$

where $\nu(\cdot) \in \mathbb{M}(\mathbb{R}_+^2 \setminus \{\mathbf{0}\})$ is called the limit or tail measure [14, 29, 44]. Using the power transformation $I \mapsto I^a$ with $a = \iota_{\text{in}}/\iota_{\text{out}}$, the vector (I^a, O) becomes standard regularly varying, i.e.,

$$t\mathbf{P}\left[\left(\frac{I^a}{t^{1/\iota_{\text{out}}}}, \frac{O}{t^{1/\iota_{\text{out}}}}\right) \in \cdot\right] \to \tilde{\nu}(\cdot), \quad \text{in } \mathbb{M}(\mathbb{R}_+^2 \setminus \{\mathbf{0}\}),$$

where $\tilde{\nu} = \nu \circ T^{-1}$ with $T(x, y) = (x^a, y)$. With this standardization, the transformed measure $\tilde{\nu}$ is directly estimable from data [50].

1.2 Contribution

For many social network data sets, degree distributions have power-law tails and one important issue is to estimate the tail exponents of these power laws so that one can predict the occurrence of nodes with large degrees. From a statistical point of view, it is useful to formulate estimation procedures using the information contained in the tails, with the hope that such an approach will be more robust against modeling errors.

In practice, power-law exponents are often estimated using an extreme value theory method relying on the *Hill* estimator. Such Hill estimates are provided as one of the key summary statistics on websites that collect large data sets from different networks, e.g. KONECT (`http://konect.cc/`). Even though standard extreme value techniques are widely adopted, no rigorous justification has been given. This suggests research questions.

In this thesis, we analyze the linear preferential attachment model from three different aspects, which leads to contribution in both statistical and probabilistic directions.

Fitting a Linear Preferential Attachment Model

This project contains two consecutive studies [63, 64] that provide useful insights on the parameter estimation problem for network data and illustrate how extreme value theory could provide robust statistical methodologies.

We develop estimation procedures for network models that exhibit power-law tail degree distributions. In [64], we propose two parametric approaches. When the full history of the network evolution is available, we derive maximum likelihood estimates (MLE) for the model parameters. If, however, only a snapshot of the nodes and edges is available at a single point in time, we give an estimation procedure for the parameters of the network using an approximation to the likelihood and method of moments. The MLE method gives strongly consistent and efficient parameter estimates for simulated data. The snapshot method leads to estimates that are consistent but with larger variances. When applied to real data, both methods display discrepancies; the MLE method does, however, suggest ways to reduce modeling error perhaps using change point methods.

Such pitfalls when calibrating the parametric model to real network data makes us look for a more robust procedure. From a practitioner's point of view, real data are never as clean as the simulated ones, so we come up with a semi-parametric estimation approach in [63], based on the estimated tail exponents and extreme value methodologies, and argue for its robustness. We find that the semi-parametric extreme value estimation method is robust in the presence of various types of data corruption.

Degree growth rates and index estimation for network data

This part is based on two papers [65, 68] and contributes to filling the gap between what is widely adopted in practice and what has shown to be true in theory. Here we only focus on the results related to the directed network model ([65]).

When analyzing network growth models, we are interested in the asymptotic behavior of the degree growth for both a fixed node and the node with maximal degree. Methodologies adopted in the literature are largely dependent on complex difference equations and martingale analyses, which are vulnerable to changes in the assumption on the particular preferential attachment rule and without a clear structural analysis, the explicit form of the asymptotic limit is not available directly from the martingale method. To remedy these drawbacks, we embed the in- and out-degree growth of a fixed node into a sequence of paired switched birth processes with immigration, using Bernoulli switching between pairs of independent birth processes with immigration. Once embedding results are available, the limiting joint distribution of the in- and out-degree of a fixed node follows by borrowing results from the asymptotics of continuous time branching processes.

In addition to the interest in the asymptotic behavior, one pragmatic issue for network data is the estimation of the power-law index. The most commonly used estimator in practice is the Hill estimator, despite the fact that its consistency is established only for data generated from repeated sampling in the extreme value literature. For the node-based network data, we view the Hill estimator as a functional of the marginal tail empirical measure. Then with the emedding technique, we prove in [65] the convergence of the tail empirical joint measure, thus giving the consistency of the two marginal Hill estimators. This allows us to reformulate known results (cf. [55, 67]) in a more structural way and close logical loops within the whole series of analyses.

Threshold Selection

When estimating the power-law index, an important step is to determine the cutoff value so that the distribution of observations larger than this threshold follows a power law. The third part of this thesis is based on [17], and studies the asymptotic behavior of the power-law tail estimates arising from a popular threshold selection method given in [10].

The minimum distance selection procedure (MDSP) given in [10] chooses the threshold as the one minimizing the Kolmogorov-Smirnov distance from the empirical tail to the estimated tail. In spite of its frequent use, no theoretical justifications for this method have been given. For instance, the consistency of the Hill estimator with the chosen threshold is not proved, even though simulation results are confirmatory. Besides, the asymptotic distribution of the chosen proportion also appears complicated but not normal. When applied to Pareto distributed data generated from simple repeated sampling, this threshold selection method tends to choose a higher threshold than the true one with non-negligible probability, thus raising questions on the trustworthiness of this widely adopted methodology. Also, the limiting distribution of the tail estimates associated with this “optimal” threshold becomes very complicated, which leads to difficulties on making inferences.

CHAPTER 2

FITTING A LINEAR PREFERENTIAL ATTACHMENT MODEL

2.1 Overview

In this chapter, we discuss methods of fitting a simple linear preferential attachment model, which is parametrized by $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}})$. The first three parameters, α, β, γ, correspond to probabilities of the 3 scenarios for adding an edge and hence sum to 1, i.e., $\alpha + \beta + \gamma = 1$. The other two, δ_{in} and δ_{out}, are tuning parameters related to growth rates. The tail indices of the marginal power laws for the in- and out-degrees can be expressed as explicit functions of $\boldsymbol{\theta}$. The graph $G(n) = (V(n), E(n))$, where $V(n)$ is the set of nodes and $E(n)$ is the set of edges at the nth iteration, evolves based on postulates that describe how new edges and nodes are formed. This construction of the network is Markov in the sense that the probabilistic rules for obtaining $G(n+1)$ once $G(n)$ is known do not require prior knowledge of earlier stages of the construction.

The Markov structure of the model allows us to construct a likelihood function based on observing $G(n_0), G(n_0+1), \ldots, G(n_0+n)$. After deriving the likelihood function, we show that it has a unique maximum at $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}})$ and that the resulting maximum likelihood estimator is strongly consistent and asymptotically normal. The normality is proved using a martingale central limit theorem applied to the score function. The limiting distribution also reveals that $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$, $\hat{\delta}_{\text{in}}$, and $\hat{\delta}_{\text{out}}$ are asymptotically independent. From these results, asymptotic properties of the MLE for the power law indices can be derived.

For some network data, only a snapshot of the nodes and edges is available at a single point in time, that is, only $G(n)$ is available for some n. In such cases, we propose an estimation procedure for the parameters of the network using an approximation to the likelihood and method of moments. This also produces strongly consistent estimators. These estimators perform reasonably well compared to the MLE where the entire evolution of the network is known but predictably there is some loss of efficiency.

We illustrate the estimation procedure for both scenarios using simulated data. Simulation plays an important role in the process of modeling networks since it provides a way to assess the performance of model fitting procedures in the idealized setting of knowing the true model. Also, after fitting a model to real data, simulation provides a check on the quality of fit. Departures from model assumptions can often be detected via simulation of multiple realizations from the fitted network. Hence it is important to have efficient simulation algorithms for producing realizations of the preferential attachment network for a given set of parameter values. We adopt a simulation method, learned from Joyjit Roy, that was inspired by [3] and is similar to that of [58]. Details on the simulation algorithm are included in Chapter A.1.

Our fitting methods are implemented in a real data setting using the Dutch Wiki talk network [42]. While one should not expect the simple 5-parameter (later extended to 7-parameter) linear preferential attachment model to fully explain a network with millions of edges, it does provide a reasonable fit to the tail behavior of the degree distributions.

2.2 Parameter estimation: MLE based on the full network history

In this chapter, we estimate the preferential attachment parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$ under two assumptions about what data is available. In the first assumption, the full evolution of the network is observed, from which the likelihood function can be computed. The resulting MLE is strongly consistent and asymptotically normal. For the second assumption, the data only consist of one snapshot of the network with n edges, without the knowledge of the network history that produced these edges. For this case we give an estimation approach through approximating the score function and moment matching, which produces parameter estimators that are also strongly consistent but less efficient than those based on the full evolution of the network. In both cases, the estimators are uniquely determined.

2.2.1 Likelihood calculation

Assume the network begins with the graph $G(n_0)$ (consisting of n_0 edges) and then evolves according to the description in Chapter 1.1.1 with parameters $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$, where $\delta_{\text{in}}, \delta_{\text{out}} > 0$ and α, β are non-negative probabilities. The γ is implicitly defined by $\gamma = 1 - \alpha - \beta$. To avoid trivial cases, we will also assume $\alpha, \beta, \gamma < 1$ for the rest of the chapter. For MLE estimation we restrict the parameter space for $\delta_{\text{in}}, \delta_{\text{out}}$ to be $[\epsilon, K]$, for some sufficiently small $\epsilon > 0$ and large K. In particular, the true value of $\delta_{\text{in}}, \delta_{\text{out}}$ is assumed to be contained in (ϵ, K). Let $e_t = (v_t, w_t)$ be the newly created edge when the random graph evolves from

$G(t-1)$ to $G(t)$. We sometimes refer to t as the time rather than the number of edges.

Assume we observe the initial graph $G(n_0)$ and the edges $\{e_t\}_{t=n_0+1}^{n}$ in the order of their formation. For $t = n_0 + 1, \ldots, n$, the values of the following variables are known:

- $N(t)$, the number of nodes in graph $G(t)$;
- $D_v^{\text{in}}(t-1)$, $D_v^{\text{out}}(t-1)$, the in- and out-degree of node v in $G(t-1)$, for all $v \in V(t-1)$;
- J_t, the scenario under which e_t is created.

Then the likelihood function is

$$\begin{aligned}
&L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}} | \, G(n_0), (e_t)_{t=n_0+1}^{n}) \\
&= \prod_{t=n_0+1}^{n} \left(\alpha \frac{D_{w_t}^{\text{in}}(t-1) + \delta_{\text{in}}}{t-1+\delta_{\text{in}} N(t-1)} \right)^{\mathbf{1}_{\{J_t=1\}}} \\
&\quad \times \prod_{t=n_0+1}^{n} \left(\beta \Big(\frac{D_{w_t}^{\text{in}}(t-1) + \delta_{\text{in}}}{t-1+\delta_{\text{in}} N(t-1)} \Big) \Big(\frac{D^{\text{out}} v_t(t-1) + \delta_{\text{out}}}{t-1+\delta_{\text{out}} N(t-1)} \Big) \right)^{\mathbf{1}_{\{J_t=2\}}} \\
&\quad \times \prod_{t=n_0+1}^{n} \left((1-\alpha-\beta) \frac{D_{v_t}^{\text{out}}(t-1) + \delta_{\text{out}}}{t-1+\delta_{\text{out}} N(t-1)} \right)^{\mathbf{1}_{\{J_t=3\}}}
\end{aligned} \tag{2.2.1}$$

and the log likelihood function is

$$
\begin{aligned}
&\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}} |\, G(n_0),(e_t)_{t=n_0+1}^n) \qquad (2.2.2)\\
&= \log\alpha \sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}} + \log\beta \sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=2\}} + \log(1-\alpha-\beta) \sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}\\
&\quad + \sum_{t=n_0+1}^n \log\left(D_{w_t}^{\text{in}}(t-1)+\delta_{\text{in}}\right)\mathbf{1}_{\{J_t\in\{1,2\}\}}\\
&\quad + \sum_{t=n_0+1}^n \log\left(D_{v_t}^{\text{out}}(t-1)+\delta_{\text{out}}\right)\mathbf{1}_{\{J_t\in\{2,3\}\}}\\
&\quad - \sum_{t=n_0+1}^n \log(t-1+\delta_{\text{in}}N(t-1))\mathbf{1}_{\{J_t\in\{1,2\}\}}\\
&\quad - \sum_{t=n_0+1}^n \log(t-1+\delta_{\text{out}}N(t-1))\mathbf{1}_{\{J_t\in\{2,3\}\}}.
\end{aligned}
$$

The score functions for $\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}}$ are calculated as follows:

$$
\begin{aligned}
&\frac{\partial}{\partial\alpha}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}} |\, G(n_0),(e_t)_{t=n_0+1}^n)\\
&\qquad = \frac{1}{\alpha}\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}} - \frac{1}{1-\alpha-\beta}\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}, \qquad (2.2.3)\\
&\frac{\partial}{\partial\beta}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}} |\, G(n_0),(e_t)_{t=n_0+1}^n)\\
&\qquad = \frac{1}{\beta}\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=2\}} - \frac{1}{1-\alpha-\beta}\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}, \qquad (2.2.4)\\
&\frac{\partial}{\partial\delta_{\text{in}}}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}} |\, G(n_0),(e_t)_{t=n_0+1}^n)\\
&\qquad = \sum_{t=n_0+1}^n \frac{1}{D_{w_t}^{\text{in}}(t-1)+\delta_{\text{in}}}\mathbf{1}_{\{J_t\in\{1,2\}\}}\\
&\qquad\quad - \sum_{t=n_0+1}^n \frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)}\mathbf{1}_{\{J_t\in\{1,2\}\}}, \qquad (2.2.5)\\
&\frac{\partial}{\partial\delta_{\text{out}}}\log L(\alpha,\beta,\delta_{\text{in}},\delta_{\text{out}} |\, G(n_0),(e_t)_{t=n_0+1}^n)\\
&\qquad = \sum_{t=n_0+1}^n \frac{1}{D_{v_t}^{\text{out}}(t-1)+\delta_{\text{out}}}\mathbf{1}_{\{J_t\in\{2,3\}\}}\\
&\qquad\quad - \sum_{t=n_0+1}^n \frac{N(t-1)}{t-1+\delta_{\text{out}}N(t-1)}\mathbf{1}_{\{J_t\in\{2,3\}\}}.
\end{aligned}
$$

Note that the score functions (2.2.3), (2.2.4) for α and β do not depend on δ_{in} and δ_{out}. One can show that the Hessian matrix of the log-likelihood for (α, β) is positive definite. Setting (2.2.3) and (2.2.4) to zero gives the unique MLE estimates for α and β,

$$\hat{\alpha}^{MLE} = \frac{1}{n - n_0} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=1\}}, \tag{2.2.6}$$

$$\hat{\beta}^{MLE} = \frac{1}{n - n_0} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=2\}}. \tag{2.2.7}$$

These estimates are strongly consistent by applying the strong law of large numbers for the $\{J_t\}$ sequence.

Next, consider the first term of the score function for δ_{in} in (2.2.5), and we have

$$\sum_{t=n_0+1}^{n} \frac{1}{D^{\text{in}}_{w_t}(t-1) + \delta_{\text{in}}} \mathbf{1}_{\{J_t \in \{1,2\}\}} = \sum_{i=0}^{\infty} \frac{1}{i + \delta_{\text{in}}} \sum_{t=n_0+1}^{n} \mathbf{1}_{\left\{D^{\text{in}}_{w_t}(t-1)=i, J_t \in \{1,2\}\right\}}.$$

Observe that $\left\{D^{\text{in}}_{w_t}(t-1) = i, J_t \in \{1,2\}\right\}$ describes the event that the in-degree of node $w_t \in V(t-1)$ is i at time $t-1$ and is augmented to $i+1$ at time t. For each $i \geq 1$, such an event happens at some stage $t \in \{n_0+1, n_0+2, \ldots, n\}$ only for those nodes with in-degree $\leq i$ at time n_0 and in-degree $> i$ at time n. Let $N_{ij}(n)$ denote the number of nodes with in-degree i and out-degree j at time n, and $N^{\text{in}}_i(n)$ and $N^{\text{in}}_{>i}(n)$ to be the number of nodes with in-degree equal to i and greater than i, respectively, i.e.,

$$N^{\text{in}}_i(n) = \sum_{j=0}^{\infty} N_{ij}(n), \quad N^{\text{in}}_{>i}(n) = \sum_{k>i} N^{\text{in}}_k(n).$$

Then

$$\sum_{t=n_0+1}^{n} \mathbf{1}_{\left\{D^{\text{in}}_{w_t}(t-1)=i, J_t \in \{1,2\}\right\}} = N^{\text{in}}_{>i}(n) - N^{\text{in}}_{>i}(n_0), \quad i \geq 1.$$

On the other hand, when $i = 0$, $\left\{D^{\text{in}}_{w_t}(t-1) = 0, J_t \in \{1,2\}\right\}$ occurs for some t if and only if all of the following three events happen:

(i) w_t has in-degree > 0 at time n;

(ii) w_t does not have in-degree > 0 at time n_0;

(iii) w_t was not created under the γ-scheme (otherwise it would have been born with in-degree 1).

This implies:

$$\sum_{t=n_0+1}^{n} \mathbf{1}_{\{D^{\text{in}}_{w_t}(t-1)=0, J_t\in\{1,2\}\}} = N^{\text{in}}_{>0}(n) - N^{\text{in}}_{>0}(n_0) - \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}},$$

since there are, in total, $\sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}}$ nodes created under the γ-scheme. Therefore,

$$\begin{aligned}
&\sum_{t=n_0+1}^{n} \frac{1}{D^{\text{in}}_{w_t}(t-1)+\delta_{\text{in}}} \mathbf{1}_{\{J_t\in\{1,2\}\}} \\
&\quad = \sum_{i=0}^{\infty} \frac{1}{i+\delta_{\text{in}}} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{D^{\text{in}}_{w_t}(t-1)=i, J_t\in\{1,2\}\}} \\
&\quad = \sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n) - N^{\text{in}}_{>i}(n_0)}{i+\delta_{\text{in}}} - \frac{\sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}}}{\delta_{\text{in}}}.
\end{aligned} \tag{2.2.8}$$

Setting the score function (2.2.5) for δ_{in} to 0 and dividing both sides by $n-n_0$ leads to

$$\begin{aligned}
&\frac{1}{n-n_0} \sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n) - N^{\text{in}}_{>i}(n_0)}{i+\delta_{\text{in}}} - \frac{1}{\delta_{\text{in}}(n-n_0)} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}} \\
&\quad - \frac{1}{n-n_0} \sum_{t=n_0+1}^{n} \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \mathbf{1}_{\{J_t\in\{1,2\}\}} \; = \; 0,
\end{aligned} \tag{2.2.9}$$

where the only unknown parameter is δ_{in}. In Chapter 2.2.2, we show that the solution to (2.2.9) actually maximizes the likelihood function in δ_{in}. Similarly, the MLE for δ_{out} can be solved from

$$\begin{aligned}
&\frac{1}{n-n_0} \sum_{j=0}^{\infty} \frac{N^{\text{out}}_{>j}(n) - N^{\text{out}}_{>j}(n_0)}{j+\delta_{\text{out}}} - \frac{\frac{1}{n-n_0} \sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=1\}}}{\delta_{\text{out}}} \\
&\quad - \frac{1}{n-n_0} \sum_{t=n_0+1}^{n} \frac{N(t-1)}{t-1+\delta_{\text{out}} N(t-1)} \mathbf{1}_{\{J_t\in\{2,3\}\}} \; = 0,
\end{aligned}$$

where $N^{\text{out}}_{>j}(n)$ is defined in the same fashion as $N^{\text{in}}_{>i}(n)$.

Remark 2.2.1. *The arguments leading to* (2.2.8) *allow us to rewrite the likelihood function* (2.2.1)*:*

$$
\begin{aligned}
& L(\alpha, \beta, \delta_{in}, \delta_{out} | \, G(n_0), (e_t)_{t=n_0+1}^n) \\
= \; & \alpha^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}}} \beta^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=2\}}} (1-\alpha-\beta)^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}} \\
& \times \prod_{t=n_0+1}^n (t-1+\delta_{in} N(t-1))^{-\mathbf{1}_{\{J_t\in\{1,2\}\}}} (t-1+\delta_{out} N(t-1))^{-\mathbf{1}_{\{J_t\in\{2,3\}\}}} \\
& \times \prod_{t=n_0+1}^n \left[\prod_{i=0}^\infty (i+\delta_{in})^{\mathbf{1}\left\{D^{in}_{w_t}(t-1)=i, J_t\in\{1,2\}\right\}} \prod_{j=0}^\infty (j+\delta_{out})^{\mathbf{1}\left\{D^{out}_{v_t}(t-1)=j, J_t\in\{2,3\}\right\}} \right] \\
= \; & \alpha^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=1\}}} \beta^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=2\}}} (1-\alpha-\beta)^{\sum_{t=n_0+1}^n \mathbf{1}_{\{J_t=3\}}} \\
& \times \prod_{t=n_0+1}^n \left[(t-1+\delta_{in} N(t-1))^{-\mathbf{1}_{\{J_t\in\{1,2\}\}}} (t-1+\delta_{out} N(t-1))^{-\mathbf{1}_{\{J_t\in\{2,3\}\}}} \delta_{in}^{-\mathbf{1}_{\{J_t=3\}}} \delta_{out}^{-\mathbf{1}_{\{J_t=1\}}} \right] \\
& \times \prod_{i=0}^\infty (i+\delta_{in})^{N^{in}_{>i}(n)-N^{in}_{>i}(n_0)} \prod_{j=0}^\infty (j+\delta_{out})^{N^{out}_{>j}(n)-N^{out}_{>j}(n_0)}.
\end{aligned}
$$

Hence by the factorization theorem, $N(n_0)$*,* $(J_t)_{t=n_0+1}^n$*,* $(N^{in}_{>i}(n) - N^{in}_{>i}(n_0))_{i\geq 0}$*,* $(N^{out}_{>j}(n) - N^{out}_{>j}(n_0))_{j\geq 0}$ *are sufficient statistics for* $(\alpha, \beta, \delta_{in}, \delta_{out})$*.*

2.2.2 Consistency of MLE

We remarked after (2.2.6) and (2.2.7) that $\hat{\alpha}^{MLE}$ and $\hat{\beta}^{MLE}$ converge almost surely to α and β. We now prove that the MLE of $(\delta_{\text{in}}, \delta_{\text{out}})$ is also strongly consistent. Note that if we initiate the network with $G(n_0)$ (for both n_0 and $N(n_0)$ finite), then almost surely for all $i, j \geq 0$,

$$
\frac{N^{\text{in}}_{>i}(n_0)}{n} \leq \frac{N(n_0)}{n} \to 0, \quad \frac{N^{\text{out}}_{>j}(n_0)}{n} \leq \frac{N(n_0)}{n} \to 0, \quad \text{as } n \to \infty,
$$

and $(n-n_0)/n \to 1$. In other words, n_0, $N^{\text{in}}_{>i}(n_0)$, $N^{\text{out}}_{>j}(n_0)$ are all $o(n)$. So for simplicity, we assume that the graph is initiated with finitely many nodes and no edges, that is, $n_0 = 0$ and $N(0) \geq 1$. In particular, these assumptions imply the sum of the in-degrees at time n is equal to n.

Let $\Psi_n(\cdot), \Phi_n(\cdot)$ be the functional forms of the terms in the log-likelihood function (2.2.2) involving δ_{in} and δ_{out} respectively, normalized by $1/n$, i.e.,

$$\begin{aligned}
\Psi_n(\lambda) &:= \sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)}{n} \log(i+\lambda) - \frac{\log\lambda}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}} \\
&\quad - \frac{1}{n} \sum_{t=1}^{n} \log\left(t-1+\lambda N(t-1)\right) \mathbf{1}_{\{J_t\in\{1,2\}\}}, \\
\Phi_n(\mu) &:= \sum_{j=0}^{\infty} \frac{N^{\text{out}}_{>j}(n)}{n} \log(j+\mu) - \frac{\log\mu}{n} \sum_{t=1}^{n} \mathbf{1}_{\{J_t=1\}} \\
&\quad - \frac{1}{n} \sum_{t=1}^{n} \log\left(t-1+\mu N(t-1)\right) \mathbf{1}_{\{J_t\in\{2,3\}\}}.
\end{aligned}$$

The following theorem gives the consistency of the MLE of δ_{in} and δ_{out}.

Theorem 2.2.2. *Suppose $\delta_{in}, \delta_{out} \in (\epsilon, K) \subset (0,\infty)$. Define*

$$\hat{\delta}^{MLE}_{in} = \hat{\delta}^{MLE}_{in}(n) := \operatorname*{argmax}_{\epsilon\leq\lambda\leq K} \Psi_n(\lambda), \quad \hat{\delta}^{MLE}_{out} = \hat{\delta}^{MLE}_{out}(n) := \operatorname*{argmax}_{\epsilon\leq\mu\leq K} \Phi_n(\mu).$$

Then these are the MLE estimators of $\delta_{in}, \delta_{out}$ and they are strongly consistent; that is,

$$\hat{\delta}^{MLE}_{in} \xrightarrow{a.s.} \delta_{in}, \qquad \hat{\delta}^{MLE}_{out} \xrightarrow{a.s.} \delta_{out}, \qquad n\to\infty.$$

Proof of Theorem 2.2.2. We only verify the consistency of $\hat{\delta}^{MLE}_{\text{in}}$ since similar arguments apply to $\hat{\delta}^{MLE}_{\text{out}}$. Define

$$\begin{aligned}
\psi_n(\lambda) := \Psi_n'(\lambda) &= \sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i+\lambda} - \frac{\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}}}{\lambda} \\
&\quad - \frac{1}{n}\sum_{t=1}^{n} \frac{N(t-1)}{t-1+\lambda N(t-1)} \mathbf{1}_{\{J_t\in\{1,2\}\}}.
\end{aligned}$$

Let us consider a limit version of ψ_n:

$$\psi(\lambda) := \sum_{i=0}^{\infty} \frac{p^{\text{in}}_{>i}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - (1-\beta)a_1(\lambda), \tag{2.2.10}$$

where $p^{\text{in}}_{>i}(\delta_{\text{in}}) := \sum_{k>i} p^{\text{in}}_k(\delta_{\text{in}})$ with $p^{\text{in}}_k(\delta_{\text{in}}) := p^{\text{in}}_k$ as defined in (1.1.4), and

$$a_1(\lambda) := \frac{\alpha+\beta}{1+\lambda(1-\beta)}, \qquad \lambda > 0.$$

Here we write $p^{\text{in}}_i(\delta_{\text{in}})$ to emphasize the dependence on δ_{in}. In Lemmas A.2.1 and A.2.2, provided in the appendix, it is shown that $\psi(\cdot)$ has a unique zero at δ_{in}, where $\psi(\lambda) > 0$ when $\lambda < \delta_{\text{in}}$ and $\psi(\lambda) < 0$ when $\lambda > \delta_{\text{in}}$, and

$$\sup_{\lambda \geq \epsilon} |\psi_n(\lambda) - \psi(\lambda)| \to 0. \tag{2.2.11}$$

Since ψ is continuous, for any $\kappa > 0$ arbitrarily small, there exists $\varepsilon_\kappa > 0$ such that $\psi(\lambda) > \varepsilon_\kappa$ for $\lambda \in [\epsilon, \delta_{\text{in}} - \kappa]$ and $\psi(\lambda) < -\varepsilon_\kappa$ for $\lambda \in [\delta_{\text{in}} + \kappa, K]$. From (2.2.11),

$$\mathbf{P}\left(\exists N_\kappa \; s.t. \sup_{n>N_\kappa} \sup_{\lambda\in[\epsilon,K]} |\psi_n(\lambda) - \psi(\lambda)| < \varepsilon_\kappa/2\right) = 1. \tag{2.2.12}$$

Note $\sup_{\lambda\in[\epsilon,K]} |\psi_n(\lambda) - \psi(\lambda)| < \varepsilon_\kappa/2$ implies

$$\psi_n(\lambda) \geq \psi(\lambda) - \sup_{\lambda\in[\epsilon,K]} |\psi_n(\lambda) - \psi(\lambda)| \geq \varepsilon_\kappa - \varepsilon_\kappa/2 > 0, \quad \lambda \in [\epsilon, \delta_{\text{in}} - \kappa),$$

and

$$\psi_n(\lambda) \leq \psi(\lambda) + \sup_{\lambda\in[\epsilon,K]} |\psi_n(\lambda) - \psi(\lambda)| \leq -\varepsilon_\kappa + \varepsilon_\kappa/2 < 0, \quad \lambda \in (\delta_{\text{in}} + \kappa, K].$$

These jointly indicate that $\delta_{\text{in}} - \kappa \leq \hat{\delta}^{MLE}_{\text{in}} \leq \delta_{\text{in}} + \kappa$. Hence (2.2.12) implies

$$\mathbf{P}\left(\lim_{n\to\infty} |\hat{\delta}^{MLE}_{\text{in}} - \delta_{\text{in}}| \leq \kappa\right) = 1,$$

for arbitrary $\kappa > 0$. That is, $\hat{\delta}^{MLE}_{\text{in}} \xrightarrow{\text{a.s.}} \delta_{\text{in}}$. □

2.2.3 Asymptotic normality of MLE

In the following theorem, we establish the asymptotic normality for the MLE estimator

$$\hat{\boldsymbol{\theta}}^{MLE}_n = (\hat{\alpha}^{MLE}, \hat{\beta}^{MLE}, \hat{\delta}^{MLE}_{\text{in}}, \hat{\delta}^{MLE}_{\text{out}}).$$

Theorem 2.2.3. *Let $\hat{\boldsymbol{\theta}}_n^{MLE}$ be the MLE estimator for $\boldsymbol{\theta}$, the parameter vector of the preferential attachment model. Then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{MLE} - \boldsymbol{\theta}) \xrightarrow{d} N\left(\mathbf{0}, \Sigma(\boldsymbol{\theta})\right),$$

where

$$\Sigma^{-1}(\boldsymbol{\theta}) = \boldsymbol{I}(\boldsymbol{\theta}) := \begin{bmatrix} \frac{1-\beta}{\alpha(1-\alpha-\beta)} & \frac{1}{1-\alpha-\beta} & 0 & 0 \\ \frac{1}{1-\alpha-\beta} & \frac{1-\alpha}{\beta(1-\alpha-\beta)} & 0 & 0 \\ 0 & 0 & I_{in} & 0 \\ 0 & 0 & 0 & I_{out} \end{bmatrix}, \tag{2.2.13}$$

with

$$\begin{aligned} I_{in} &:= \sum_{i=0}^{\infty} \frac{p_{>i}^{in}}{(i+\delta_{in})^2} - \frac{\gamma}{\delta_{in}^2} - \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{in}(1-\beta))^2}, \\ I_{out} &:= \sum_{j=0}^{\infty} \frac{p_{>j}^{out}}{(j+\delta_{out})^2} - \frac{\alpha}{\delta_{out}^2} - \frac{(\gamma+\beta)(1-\beta)^2}{(1+\delta_{out}(1-\beta))^2}. \end{aligned} \tag{2.2.14}$$

In particular, $\boldsymbol{I}(\boldsymbol{\theta})$ is the asymptotic Fisher information matrix for the parameters, and hence the MLE estimator is efficient.

Remark 2.2.4. *From Theorem 2.2.3, the estimators $(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE})$, $\hat{\delta}_{in}^{MLE}$, and $\hat{\delta}_{out}^{MLE}$ are asymptotically independent.*

Proof of Theorem 2.2.3. We first show the limiting distributions for the MLE's, i.e. $(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE})$, $\hat{\delta}_{\mathrm{in}}^{MLE}$ and $\hat{\delta}_{\mathrm{out}}^{MLE}$. From (2.2.6) and (2.2.7),

$$(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE}) = \frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{1}_{\{J_t=1\}}, \mathbf{1}_{\{J_t=2\}}\right),$$

where $\{J_t\}$ is a sequence of iid random variables. Hence the limiting distribution of the pair $\left(\hat{\alpha}^{MLE}, \hat{\beta}^{MLE}\right)$ follows directly from standard central limit theorem for sums of independent random variables.

Next we show the asymptotic normality for $\hat{\delta}_{\text{in}}^{MLE}$; the argument for $\hat{\delta}_{\text{out}}^{MLE}$ is similar. Recall from (2.2.5) that the score function for δ_{in} can be written as

$$\frac{\partial}{\partial \delta_{\text{in}}} \log L(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})\bigg|_{\delta} =: \sum_{t=1}^{n} u_t(\delta),$$

where u_t is defined by

$$u_t(\delta) := \frac{1}{D_{(t-1)}^{\text{in}}(w_t) + \delta} \mathbf{1}_{\{J_t \in \{1,2\}\}} - \frac{N(t-1)}{t-1+\delta N(t-1)} \mathbf{1}_{\{J_t \in \{1,2\}\}}. \tag{2.2.15}$$

The MLE estimator $\hat{\delta}_{\text{in}}^{MLE}$ can be obtained by solving $\sum_{t=1}^{n} u_t(\delta) = 0$. By a Taylor expansion of $\sum_{t=1}^{n} u_t(\delta)$,

$$0 = \sum_{t=1}^{n} u_t(\hat{\delta}_{\text{in}}^{MLE}) = \sum_{t=1}^{n} u_t(\delta_{\text{in}}) + (\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}) \sum_{t=1}^{n} \dot{u}_t(\hat{\delta}_{\text{in}}^{*}), \tag{2.2.16}$$

where $\dot{u}_t$ denotes the derivative of u_t and $\hat{\delta}_{\text{in}}^{*} = \delta_{\text{in}} + \xi(\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}})$ for some $\xi \in [0,1]$. An elementary transformation of (2.2.16) gives

$$n^{1/2}(\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}) = \left(-\frac{1}{n^{-1}\sum_{t=1}^{n} \dot{u}_t(\hat{\delta}_{\text{in}}^{*})}\right)\left(n^{-1/2}\sum_{t=1}^{n} u_t(\delta_{\text{in}})\right).$$

To establish

$$n^{1/2}(\hat{\delta}_{\text{in}}^{MLE} - \delta_{\text{in}}) \xrightarrow{d} N(0, I_{\text{in}}^{-1}),$$

where I_{in} is as defined in (2.2.13), it suffices to show the following two results:

(i) $n^{-1/2}\sum_{t=1}^{n} u_t(\delta_{\text{in}}) \xrightarrow{d} N(0, I_{\text{in}})$,

(ii) $n^{-1}\sum_{t=1}^{n} \dot{u}_t(\hat{\delta}_{\text{in}}^{*}) \xrightarrow{p} -I_{\text{in}}$.

These are proved in Lemmas A.3.1 and A.3.2 in the appendix, respectively.

To establish the joint asymptotic normality of the MLE estimator $\hat{\boldsymbol{\theta}}_n^{MLE}$, denote the joint score function vector for $\boldsymbol{\theta}$ by

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta}) =: \mathbf{S}_n(\boldsymbol{\theta}) = (S_n(\alpha), S_n(\beta), S_n(\delta_{\text{in}}), S_n(\delta_{\text{out}}))^T,$$

where $S_n(\alpha), S_n(\beta), S_n(\delta_{\text{in}}), S_n(\delta_{\text{out}})$ are the score functions for $\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}$, respectively. A multivariate Taylor expansion gives

$$\mathbf{0} \; = \; \mathbf{S}_n\left(\hat{\boldsymbol{\theta}}_n^{MLE}\right) \; = \; \mathbf{S}_n(\boldsymbol{\theta}) + \dot{\mathbf{S}}_n\left(\hat{\boldsymbol{\theta}}_n^{*}\right)\left(\hat{\boldsymbol{\theta}}_n^{MLE} - \boldsymbol{\theta}\right), \tag{2.2.17}$$

where $\dot{\mathbf{S}}_n$ denotes the Hessian matrix of the log-likelihood function $\log L(\boldsymbol{\theta})$, and $\hat{\boldsymbol{\theta}}_n^{*} = \boldsymbol{\theta} + \boldsymbol{\xi} \circ \left(\hat{\boldsymbol{\theta}}_n^{MLE} - \boldsymbol{\theta}\right)$ for some vector $\boldsymbol{\xi} \in [0,1]^4$, where "$\circ$" denotes the Hadamard product. From Remark 2.2.1, the likelihood function $L(\boldsymbol{\theta})$ can be factored into

$$L(\boldsymbol{\theta}) \; = \; f_1(\alpha, \beta) f_2(\delta_{\text{in}}) f_3(\delta_{\text{out}}).$$

Hence

$$\frac{1}{n}\dot{\mathbf{S}}_n(\hat{\boldsymbol{\theta}}_n^{*}) \; = \; \begin{bmatrix} \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \alpha^2} & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \alpha \partial \beta} & 0 & 0 \\ \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \beta \partial \alpha} & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \beta^2} & 0 & 0 \\ 0 & 0 & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \delta_{\text{in}}^2} & 0 \\ 0 & 0 & 0 & \frac{\partial^2 \log L_n(\hat{\boldsymbol{\theta}}_n^{*})}{\partial \delta_{\text{out}}^2} \end{bmatrix} \stackrel{p}{\to} I(\boldsymbol{\theta}) \tag{2.2.18}$$

as implied in the previous part of the proof, where $I(\boldsymbol{\theta})$ (defined in (2.2.13)) is positive semi-definite.

Note that $(S_n(\alpha), S_n(\beta)), S_n(\delta_{\text{in}}), S_n(\delta_{\text{out}})$ are pairwise uncorrelated. As an example, observe that

$$\begin{aligned} \mathbf{E}[S_n(\alpha) S_n(\delta_{\text{in}})] \; &= \; \int \frac{\partial \log L(\boldsymbol{\theta})}{\partial \alpha} \frac{\partial \log L(\boldsymbol{\theta})}{\partial \delta_{\text{in}}} L(\boldsymbol{\theta}) d\mathbf{x} \\ &= \; \int \frac{\partial \log f_1(\alpha,\beta)}{\partial \alpha} \frac{\partial \log f_2(\delta_{\text{in}})}{\partial \delta_{\text{in}}} f_1(\alpha,\beta) f_2(\delta_{\text{in}}) f_3(\delta_{\text{out}}) d\mathbf{x} \\ &= \; \int \frac{\partial f_1(\alpha,\beta)}{\partial \alpha} \frac{\partial f_2(\delta_{\text{in}})}{\partial \delta_{\text{in}}} f_3(\delta_{\text{out}})\, d\mathbf{x} \\ &= \; \frac{\partial^2}{\partial \alpha \partial \delta_{\text{in}}} \int L(\boldsymbol{\theta}) d\mathbf{x} \\ &= \; 0 \; = \; \mathbf{E}[S_n(\alpha)]\mathbf{E}[S_n(\delta_{\text{in}})]. \end{aligned}$$

Using the Cramér-Wold device, the joint convergence of $\mathbf{S}_n(\boldsymbol{\theta})$ follows easily, i.e.,

$$n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}) \; \stackrel{d}{\to} \; N(\mathbf{0}, I(\boldsymbol{\theta})).$$

From here, the result of the theorem follows from (2.2.17) and (2.2.18). $\square$

2.3 Parameter estimation based on one snapshot

Based only on the single snapshot $G(n)$, we propose a parameter estimation procedure. We assume that the choice of the snapshot does not depend on any endogenous information related to the network. The snapshot merely represents a point in time where the data is available. Since no information on the initial graph $G(n_0)$ is available, we merely assume n_0 and $N(n_0)$ are fixed and $n \to \infty$.

Among the sufficient statistics for $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$ derived in Remark 2.2.1, $\left(N^{\text{in}}_{>i}(n)\right)_{i\geq 0}$, $\left(N^{\text{out}}_{>j}(n)\right)_{j\geq 0}$ are computable from $G(n)$, but the $(J_t)_{t=1}^n$ are not. However, when n is large, we can use the following approximations according to the proof of Lemma A.2.2:

$$\frac{1}{n}\sum_{t=n_0+1}^{n} \mathbf{1}_{\{J_t=3\}} \;\approx\; 1-\alpha-\beta,$$

and

$$\frac{1}{n}\sum_{t=n_0+1}^{n} \frac{N(t)}{t+\delta_{\text{in}}N(t)}\mathbf{1}_{\{J_t\in\{1,2\}\}} \;\approx\; (\alpha+\beta)\frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)}.$$

Substituting in (2.2.9), we estimate δ_{in} in terms of α and β by solving

$$\sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i+\delta_{\text{in}}} - \frac{1-\alpha-\beta}{\delta_{\text{in}}} - \frac{(\alpha+\beta)(1-\beta)}{1+(1-\beta)\delta_{\text{in}}} \;=\; 0. \tag{2.3.1}$$

Note that a strongly consistent estimator of β can be obtained directly from $G(n)$:

$$\tilde{\beta} \;=\; 1-\frac{N(n)}{n} \xrightarrow{\text{a.s.}} \beta.$$

To obtain an estimate for α, we make use of the recursive formula for $\{p^{\text{in}}_i\}$ in (A.2.1a):

$$\left(1+\frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}\right)p^{\text{in}}_0 \;=\; \alpha, \tag{2.3.2}$$

and replace p_0^{in} by $N_0^{\text{in}}(n)/n$ for large n,

$$\left(1+\frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}\right)\frac{N_0^{\text{in}}(n)}{n} = \alpha. \tag{2.3.3}$$

Plug the strongly consistent estimator $\tilde{\beta}$ into (2.3.1) and (2.3.3), and we claim that solving the system of equations:

$$\sum_{i=0}^{\infty}\frac{N_{>i}^{\text{in}}(n)/n}{i+\delta_{\text{in}}}-\frac{1-\alpha-\tilde{\beta}}{\delta_{\text{in}}}-\frac{(\alpha+\tilde{\beta})(1-\tilde{\beta})}{1+(1-\tilde{\beta})\delta_{\text{in}}} = 0, \tag{2.3.4a}$$

$$\left(1+\frac{(\alpha+\tilde{\beta})\delta_{\text{in}}}{1+(1-\tilde{\beta})\delta_{\text{in}}}\right)\frac{N_0^{\text{in}}(n)}{n} = \alpha, \tag{2.3.4b}$$

gives the unique solution $(\tilde{\alpha},\tilde{\delta}_{\text{in}})$ which is strongly consistent for $(\alpha,\delta_{\text{in}})$.

Theorem 2.3.1. *The solution $(\tilde{\alpha},\tilde{\delta}_{in})$ to the system of equations in* (2.3.4) *is unique and strongly consistent for (α,δ_{in}), i.e.*

$$\tilde{\alpha}\xrightarrow{a.s.}\alpha,\quad \tilde{\delta}_{in}\xrightarrow{a.s.}\delta_{in}.$$

The proof of Theorem 2.3.1 is given in Chapter A.4.

The parameters $\tilde{\delta}_{\text{out}}$ and $\tilde{\gamma}$ can be estimated by a mirror argument. We summarize the estimation procedure for $(\alpha,\beta,\gamma,\delta_{\text{in}},\delta_{\text{out}})$ from the snapshot $G(n)$ as follows:

1. Estimate β by $\tilde{\beta} = 1-N(n)/n$.
2. Obtain $\tilde{\delta}_{\text{in}}^0$ by solving (i.e., matching (2.3.4a) and (2.3.4b))

$$\sum_{i=1}^{\infty}\frac{N_{>i}^{\text{in}}(n)}{n}\frac{i}{i+\delta_{\text{in}}}(1+\delta_{\text{in}}(1-\tilde{\beta})) = \frac{\frac{N_0^{\text{in}}(n)}{n}+\tilde{\beta}}{1-\frac{N_0^{\text{in}}(n)}{n}\frac{\delta_{\text{in}}}{1+(1-\tilde{\beta})\delta_{\text{in}}}}.$$

3. Estimate α by

$$\tilde{\alpha}^0 = \frac{\frac{N_0^{\text{in}}(n)}{n}+\tilde{\beta}}{1-\frac{N_0^{\text{in}}(n)}{n}\frac{\tilde{\delta}_{\text{in}}^0}{1+(1-\tilde{\beta})\tilde{\delta}_{\text{in}}^0}}-\tilde{\beta}.$$

4. Obtain $\tilde{\delta}^0_{\text{out}}$ by solving

$$\sum_{j=1}^{\infty} \frac{N^{\text{out}}_{>j}(n)}{n} \frac{j}{j+\delta_{\text{out}}}(1+\delta_{\text{out}}(1-\tilde{\beta})) = \frac{\frac{N^{\text{out}}_0(n)}{n} + \tilde{\beta}}{1 - \frac{N^{\text{out}}_0(n)}{n} \frac{\delta_{\text{out}}}{1+(1-\tilde{\beta})\delta_{\text{out}}}}.$$

5. Estimate γ by

$$\tilde{\gamma}^0 = \frac{\frac{N^{\text{out}}_0(n)}{n} + \tilde{\beta}}{1 - \frac{N^{\text{out}}_0(n)}{n} \frac{\tilde{\delta}^0_{\text{out}}}{1+(1-\tilde{\beta})\tilde{\delta}^0_{\text{out}}}} - \tilde{\beta}.$$

Note that even though all three estimators $\tilde{\alpha}^0, \tilde{\beta}, \tilde{\gamma}^0$ are strongly consistent and hence $\tilde{\alpha}^0 + \tilde{\beta} + \tilde{\gamma}^0 \xrightarrow{\text{a.s.}} 1$, Step 1–5 do not necessarily imply the equality

$$\tilde{\alpha}^0 + \tilde{\beta} + \tilde{\gamma}^0 = 1.$$

We recommend adding the following two steps for a re-normalization to overcome this defect.

6. Re-normalize the probabilities

$$(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \leftarrow \left(\frac{\tilde{\alpha}^0(1-\tilde{\beta})}{\tilde{\alpha}^0 + \tilde{\gamma}^0}, \tilde{\beta}, \frac{\tilde{\gamma}^0(1-\tilde{\beta})}{\tilde{\alpha}^0 + \tilde{\gamma}^0} \right).$$

7. Plug $\tilde{\alpha}$ into (2.3.4a) to update the estimate of δ_{in}, i.e., solve for $\tilde{\delta}_{\text{in}}$ from

$$\sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i+\tilde{\delta}_{\text{in}}} - \frac{1-\tilde{\alpha}-\tilde{\beta}}{\tilde{\delta}_{\text{in}}} - \frac{(\tilde{\alpha}+\tilde{\beta})(1-\tilde{\beta})}{1+(1-\tilde{\beta})\tilde{\delta}_{\text{in}}} = 0.$$

   Similarly, solve for $\tilde{\delta}_{\text{out}}$ from

$$\sum_{j=0}^{\infty} \frac{N^{\text{out}}_{>j}(n)/n}{j+\tilde{\delta}_{\text{out}}} - \frac{1-\tilde{\gamma}-\tilde{\beta}}{\tilde{\delta}_{\text{out}}} - \frac{(\tilde{\gamma}+\tilde{\beta})(1-\tilde{\beta})}{1+(1-\tilde{\beta})\tilde{\delta}_{\text{out}}} = 0.$$

2.4 Simulation study

We now apply the estimation procedures described in Chapter 2.2 and 2.3 to simulated data, which allows us to compare the estimation results using the full history of the network with that using just one snapshot. Algorithm 1 is used to simulate realizations of the preferential attachment network.

2.4.1 MLE

For the case of observing the full history of the network, we simulated 5000 independent replications of the preferential attachment network with 10^5 edges under the true parameter values

$$\boldsymbol{\theta} = (\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}) = (0.3,\ 0.5,\ 2,\ 1). \tag{2.4.1}$$

For each realization, the MLE estimate of the parameters was computed and standardized as

$$\frac{\sqrt{n}\left((\hat{\boldsymbol{\theta}}_n^{MLE})_i - (\boldsymbol{\theta})_i\right)}{\hat{\sigma}_{ii}}, \tag{2.4.2}$$

where $(\hat{\boldsymbol{\theta}}_n)_i$ and $(\boldsymbol{\theta})_i$ denote the i-th components of $\hat{\boldsymbol{\theta}}_n^{MLE}$ and $\boldsymbol{\theta}$ respectively, and $\hat{\sigma}_{ii}^2$ is the i-th diagonal component of the matrix $\hat{\Sigma} := \Sigma(\hat{\boldsymbol{\theta}}_n^{MLE})$. The explicit formula for the entries of $\hat{\Sigma}$ is

$$\hat{\Sigma} = \begin{bmatrix} \hat{\alpha}^{MLE}\left(1 - \hat{\alpha}^{MLE}\right) & -\hat{\alpha}^{MLE}\hat{\beta}^{MLE} & 0 & 0 \\ -\hat{\alpha}^{MLE}\hat{\beta}^{MLE} & \hat{\beta}^{MLE}\left(1 - \hat{\beta}^{MLE}\right) & 0 & 0 \\ 0 & 0 & \hat{I}_{\text{in}}^{-1} & 0 \\ 0 & 0 & 0 & \hat{I}_{\text{out}}^{-1} \end{bmatrix},$$

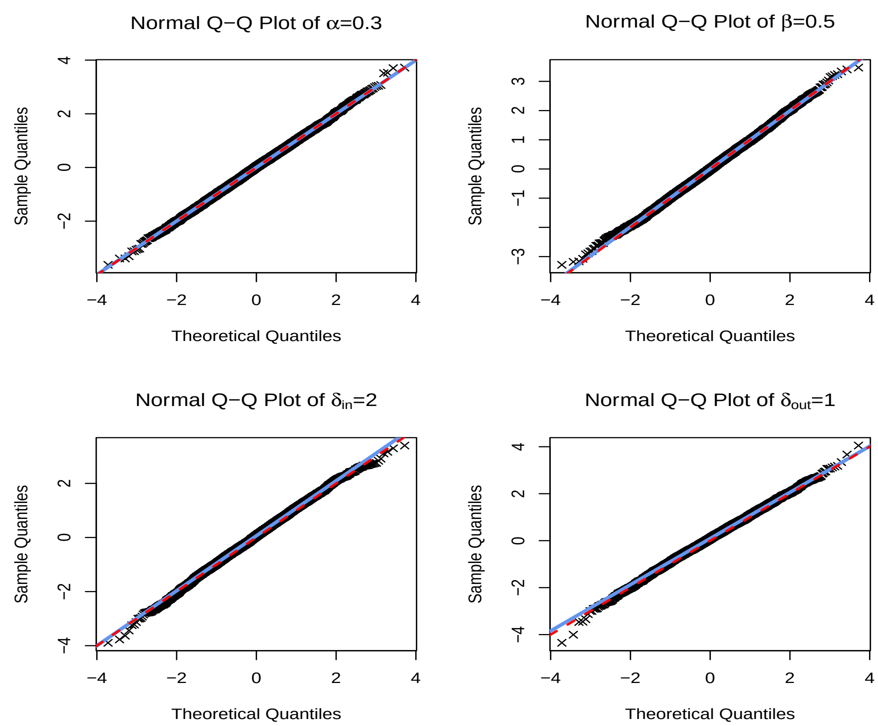


Figure 2.4.1: Normal QQ-plots in black for normalized estimates in (2.4.2) under 5000 replications of a preferential attachment network with 10^5 edges and $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$. The fitted lines in blue are the traditional qq-lines (given by R) used to check normality of the estimates. The red dashed line represents the $y = x$ line in all plots.

where, see (2.2.13) and (2.2.14),

$$\begin{aligned}
\hat{I}_{\text{in}} &= \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n}{\left(i + \hat{\delta}_{\text{in}}^{MLE}\right)^2} - \frac{1 - \hat{\alpha}^{MLE} - \hat{\beta}^{MLE}}{\left(\hat{\delta}_{\text{in}}^{MLE}\right)^2} \\
&\qquad - \frac{\left(\hat{\alpha}^{MLE} + \hat{\beta}^{MLE}\right)\left(1 - \hat{\beta}^{MLE}\right)^2}{\left(1 + \hat{\delta}_{\text{in}}^{MLE}(1 - \hat{\beta}^{MLE})\right)^2}, \\
\hat{I}_{\text{out}} &= \sum_{j=0}^{\infty} \frac{N_{>j}^{\text{out}}(n)/n}{\left(j + \hat{\delta}_{\text{out}}^{MLE}\right)^2} - \frac{\hat{\alpha}^{MLE}}{\left(\hat{\delta}_{\text{out}}^{MLE}\right)^2} - \frac{\left(1 - \hat{\alpha}^{MLE}\right)\left(1 - \hat{\beta}^{MLE}\right)^2}{\left(1 + \hat{\delta}_{\text{out}}^{MLE}(1 - \hat{\beta}^{MLE})\right)^2}.
\end{aligned}$$

By the strong consistency of the MLEs combined with Lemma A.2.2, we have that $\hat{\Sigma} \xrightarrow{\text{a.s.}} \Sigma$.

The QQ-plots of the normalized MLEs are shown in Figure 2.4.1, all of which line up quite well with the $y = x$ line (the red dashed line). This is consistent with the asymptotic theory described in Theorem 2.2.3. Confidence intervals for $\boldsymbol{\theta}$ can be obtained using this theorem. Given a single realization, an approximate $(1 - \varepsilon)$-confidence interval for $(\boldsymbol{\theta})_i$ is

$$(\hat{\boldsymbol{\theta}}_n^{MLE})_i \pm z_{\varepsilon/2}\sqrt{\frac{\hat{\sigma}_{ii}^2}{n}} \quad \text{for } i = 1, \ldots, 4,$$

where $z_{\varepsilon/2}$ is the upper $\varepsilon/2$ quantile of $N(0,1)$.

2.4.2 One snapshot

We used the same simulated data as in Chapter 2.4.1 to obtain parameter estimates $\tilde{\boldsymbol{\theta}}_n := (\tilde{\alpha}, \tilde{\beta}, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}})$ through only the final snapshot, i.e., the set of directed edges without timestamps, following the procedure described at the end of Chapter 2.3. For the purpose of comparison with MLE, Figure 2.4.2 gives the QQ-plots for the normalized estimates from the snapshots using the same standardizations for the MLEs, i.e.,

$$\frac{\sqrt{n}\left((\tilde{\boldsymbol{\theta}}_n)_i - (\boldsymbol{\theta})_i\right)}{\hat{\sigma}_{ii}}, \quad i = 1, \ldots, 4, \tag{2.4.3}$$

where $(\tilde{\boldsymbol{\theta}}_n)_i$ denotes the i-th components of $\tilde{\boldsymbol{\theta}}_n$. Again, the fitted lines in blue are the traditional QQ-lines and the red dashed lines are the $y = x$ line. The QQ-plot for $\tilde{\beta}$ exhibits the same shape as for $\hat{\beta}^{MLE}$, since the two estimates are identical.

From Figure 2.4.2, we see that the snapshot estimates of all four parameters are consistent and approximately normal, i.e., the QQ-plots are linear. However, the slopes of the QQ-lines for $\tilde{\alpha}, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}}$ are much steeper than the diagonal line,

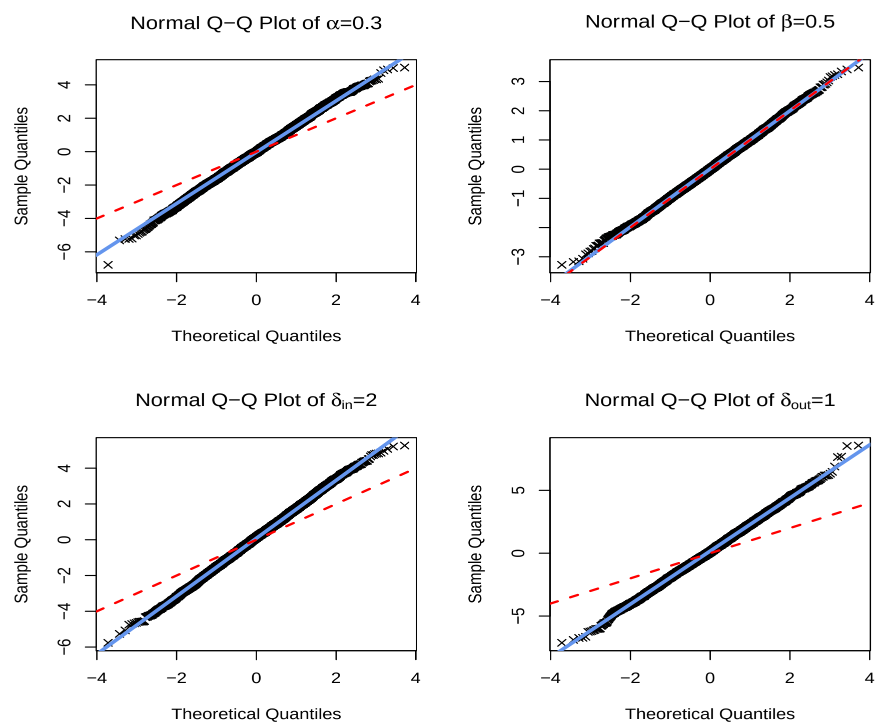


Figure 2.4.2: Normal QQ-plots for the normalized estimates in (2.4.3) under 5000 replications of a preferential attachment network with 10^5 edges and $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$. The fitted lines in blue are the traditional qq-lines used to check normality of the estimates. The red dashed line represents the $y = x$ line in all plots.

indicating a loss of efficiency for $\tilde{\boldsymbol{\theta}}_n$ compared with $\hat{\boldsymbol{\theta}}_n$. Indeed the estimator variance is inflated for all parameters except for β, where $\tilde{\beta}$ coincides with the true MLE. This is as expected since knowing only the final snapshot provides far less information than the whole network history.

Recall that for a consistent estimator T_n of a one-dimensional parameter θ constructed from a random sample of size n, the asymptotic relative efficiency (ARE) of T_n is defined by

$$ARE(T_n) := \lim_{n\to\infty} \frac{\mathrm{Var}(\sqrt{n}T_n^*)}{\mathrm{Var}(\sqrt{n}T_n)},$$

where T_n^* denotes the asymptotically efficient estimator. We may compute the ARE's for the snapshot parameter estimates

$$ARE(\tilde{\alpha}) = \lim_{n\to\infty} \frac{n\text{Var}(\hat{\alpha}^{MLE})}{n\text{Var}(\tilde{\alpha})} \approx \frac{\widehat{\text{Var}}(\hat{\alpha}^{MLE})}{\widehat{\text{Var}}(\tilde{\alpha})} \approx 0.398,$$

$$ARE(\tilde{\delta}_{\text{in}}) = \lim_{n\to\infty} \frac{n\text{Var}(\hat{\delta}_{\text{in}}^{MLE})}{n\text{Var}(\tilde{\delta}_{\text{in}})} \approx \frac{\widehat{\text{Var}}(\hat{\delta}_{\text{in}}^{MLE})}{\widehat{\text{Var}}(\tilde{\delta}_{\text{in}})} \approx 0.392,$$

$$ARE(\tilde{\delta}_{\text{out}}) = \lim_{n\to\infty} \frac{n\text{Var}(\hat{\delta}_{\text{out}}^{MLE})}{n\text{Var}(\tilde{\delta}_{\text{out}})} \approx \frac{\widehat{\text{Var}}(\hat{\delta}_{\text{out}}^{MLE})}{\widehat{\text{Var}}(\tilde{\delta}_{\text{out}})} \approx 0.226,$$

where $\widehat{\text{Var}}$ denotes the sample variance of the parameter estimate based on the 5000 replications. Note that $ARE(\tilde{\beta}) = 1$ since $\tilde{\beta} = \hat{\beta}^{MLE}$.

Given a single realization, the variances of the snapshot estimates can be estimated through resampling as follows. Using the estimated parameter $\tilde{\boldsymbol{\theta}}_n$, simulate 10^4 independent bootstrap replicates of the network with $n = 10^5$ edges. For each simulated network, the snapshot estimate, $\tilde{\boldsymbol{\theta}}_n^* := \left(\tilde{\alpha}^*, \tilde{\beta}^*, \tilde{\delta}_{\text{in}}^*, \tilde{\delta}_{\text{out}}^*\right)$, is computed. The sample variance of these 10^4 snapshot estimates can then be used as an approximation for the variance of $\tilde{\boldsymbol{\theta}}_n$ so that assuming asymptotic normality, a $(1-\varepsilon)$-confidence interval for $\boldsymbol{\theta}$ can be approximated by

$$(\tilde{\boldsymbol{\theta}}_n)_i \pm z_{\varepsilon/2} \sqrt{\widehat{\text{Var}}\left((\tilde{\boldsymbol{\theta}}_n^*)_i\right)} \quad \text{for } i = 1, \ldots, 4,$$

where $z_{\varepsilon/2}$ is the upper $\varepsilon/2$ quantile of $N(0,1)$.

2.4.3 Sensitivity test

Now we investigate the sensitivity of our estimates while values of the parameters $(n, \alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}})$ are allowed to vary. First consider the impact of n, the number of edges in the network. To do so we held the parameters fixed with values given by (2.4.1): $(\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}) = (0.3,\ 0.5,\ 2,\ 1)$ and varied the value of n. The

Table 2.4.1: Mean of $\hat{\boldsymbol{\theta}}_n^{MLE}$ and $\tilde{\boldsymbol{\theta}}_n$ with ARE's of $\tilde{\boldsymbol{\theta}}_n$ relative to $\hat{\boldsymbol{\theta}}_n^{MLE}$ for $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$ under different choices of n.

n	$Mean(\hat{\boldsymbol{\theta}}_n^{MLE})$	$Mean(\tilde{\boldsymbol{\theta}}_n)$	$ARE(\tilde{\boldsymbol{\theta}}_n)$
1000	(0.300, 0.500, 2.076, 1.054)	(0.301, 0.500, 2.128, 1.066)	(0.408, 1.000, 0.397, 0.228)
5000	(0.300, 0.500, 2.022, 1.013)	(0.301, 0.500, 2.036, 1.010)	(0.414, 1.000, 0.386, 0.236)
10000	(0.300, 0.500, 2.011, 1.006)	(0.301, 0.500, 2.019, 1.006)	(0.408, 1.000, 0.388, 0.232)
50000	(0.300, 0.500, 2.003, 1.002)	(0.300, 0.500, 2.005, 1.002)	(0.399, 1.000, 0.393, 0.230)
100000	(0.300, 0.500, 2.001, 1.001)	(0.300, 0.500, 2.003, 1.000)	(0.392, 1.000, 0.382, 0.223)

QQ-plots (not presented) for standardized estimates using both full MLE and one-snapshot methods were produced to check the asymptotic normality. When $n = 500, 1000$, diagnostics revealed departures from normality for both the MLE and the snapshot estimates. However, after increasing n to 10000, estimates obtained from both approaches appeared normally distributed as expected.

For each value of n in Table 2.4.1, 5000 replicates of the network with n edges and parameters $\boldsymbol{\theta} = (0.3, 0.5, 2, 1)$ were generated. For each realization, the MLE's $\hat{\boldsymbol{\theta}}_n^{MLE}$ were computed using the full history of the network and the one-snapshot estimates $\tilde{\boldsymbol{\theta}}_n$ were obtained using the 7-step snapshot method proposed in Chapter 2.3, pretending that only the last snapshot $G(n)$ was available. The mean for these two estimators were recorded in Table 2.4.1. There is little bias for both estimates of α and β, even for small values of n. On the other hand, there is some bias for estimated δ_{in} and δ_{out} for $n \leq 5000$. The magnitude of the biases for both types of estimates decrease as n increases. Also the ARE's of the snapshot estimator stay within a narrow band as n increases.

Next we held $(n, \delta_{\text{in}}, \delta_{\text{out}}) = (10^5, 2, 1)$ fixed and experimented with various values of (α, β) in Table 2.4.2. For each choice of (α, β), 5000 independent realizations of the network were generated and the means of the MLE $\hat{\boldsymbol{\theta}}_n^{MLE}$ and the one-snapshot estimates $\tilde{\boldsymbol{\theta}}_n$ were recorded. Overall, the biases for $\hat{\boldsymbol{\theta}}_n^{MLE}$ are remarkably small for virtually all combinations of parameter values, except for those parameter choices where one of (α, β) is extremely small. The biases for the snapshot estimates $\tilde{\boldsymbol{\theta}}_n$ exhibit a similar property, but the magnitudes of the biases are consistently larger than those in the MLE case.

In general, the snapshot estimators are able to achieve 20%–50% efficiency over the range of parameters considered. The loss of efficiency might be less than one would expect given the substantial reduction in the data available to produce the snapshot estimates. It is worth noting that in the case where $(\alpha, \beta) = (0.7, 0.2)$, the efficiencies of the snapshot estimators for α and δ_{in} are much larger (0.73 and 0.79, respectively). A heuristic explanation for this increase is that the parameter $\gamma = 1 - \alpha - \beta = 0.1$ is relatively small. By the implicit constraints used for the snapshot estimates, we have

$$\tilde{\alpha} + \tilde{\gamma} \;=\; 1 - \tilde{\beta} \;=\; 1 - \hat{\beta}^{MLE} \;=\; \hat{\alpha}^{MLE} + \hat{\gamma}^{MLE},$$

that is, the snapshot estimate of the sum $\alpha + \gamma$ is the same as the MLE for the sum. Now if γ is small, one would expect the resulting estimates to also be small so that $\tilde{\alpha}$ would be nearly the same as $\hat{\alpha}^{MLE}$. Hence the ARE would be close to 1. On the other hand, in the case of a larger γ, see the bottom row of Table 2.4.2 in which $\gamma = 0.6$, the ARE for α is not as large (0.42), but the ARE for $\tilde{\delta}_{\text{out}}$ is (0.63).

Table 2.4.2: Mean of $\hat{\boldsymbol{\theta}}_n^{MLE}$ and $\tilde{\boldsymbol{\theta}}_n$ with ARE's of $\tilde{\boldsymbol{\theta}}_n$ relative to $\hat{\boldsymbol{\theta}}_n^{MLE}$ for $(n, \delta_{\text{in}}, \delta_{\text{out}}) = (10^5, 2, 1)$ under different choices of (α, β).

(α, β)	$Mean(\hat{\boldsymbol{\theta}}_n^{MLE})$	$Mean(\tilde{\boldsymbol{\theta}}_n)$	$ARE(\tilde{\boldsymbol{\theta}}_n)$
(0.001, 0.99)	(0.001, 0.990, 2.034, 1.016)	(0.001, 0.990, 2.071, 1.049)	(0.291, 1.000, 0.147, 0.316)
(0.01, 0.9)	(0.010, 0.900, 2.004, 1.001)	(0.010, 0.900, 2.008, 1.004)	(0.331, 1.000, 0.207, 0.381)
(0.1, 0.8)	(0.100, 0.800, 2.003, 1.001)	(0.100, 0.800, 2.004, 1.002)	(0.353, 1.000. 0.264, 0.216)
(0.2, 0.6)	(0.200, 0.600, 2.002, 1.001)	(0.200, 0.600, 2.003, 1.001)	(0.364, 1.000, 0.309, 0.236)
(0.5, 0.3)	(0.500, 0.300, 2.001, 1.001)	(0.500, 0.300, 2.002, 1.000)	(0.472, 1.000, 0.529, 0.202)
(0.7, 0.2)	(0.700, 0.200, 2.002, 1.000)	(0.700, 0.200, 2.002, 1.000)	(0.726, 1.000, 0.793, 0.217)
(0.1, 0.3)	(0.100, 0.300, 2.001, 1.001)	(0.100, 0.300, 2.002, 1.000)	(0.420, 1.000, 0.313, 0.629)

2.5 Real network example

In this section, we explore fitting a preferential attachment model to a social network. As illustration, we chose the Dutch Wiki talk network dataset, available on KONECT [42] (`http://konect.uni-koblenz.de/networks/wiki_talk_nl`). The nodes represent users of Dutch Wikipedia, and an edge from node A to node B refers to user A writing a message on the talk page of user B at a certain time point. The network consists of 225,749 nodes (users) and 1,554,699 edges (messages). All edges are recorded with timestamps.

In order to accommodate all the edge formulation scenarios appeared in the dataset, we extend our model by appending the following two interaction schemes ($J_n = 4, 5$) in addition to the existing three ($J_n = 1, 2, 3$) described in Chapter 1.1.1.

- If $J_n = 4$ (with probability ξ), append to $G(n-1)$ two new nodes $v, w \in V(n) \setminus V(n-1)$ and an edge connecting them (v, w).
- If $J_n = 5$ (with probability ρ), append to $G(n-1)$ a new node $v \in V(n) \setminus V(n-1)$

with self loop (v, v).

These scenarios have been observed in other social network data, such as the network that models Facebook wall posts (`http://konect.uni-koblenz.de/networks/facebook-wosn-wall`). They occur in small proportions and can be easily accommodated by a slight modification in the model fitting procedure. The new model has parameter vector $(\alpha, \beta, \gamma, \xi, \delta_{\text{in}}, \delta_{\text{out}})$, and ρ is implicitly defined through $\rho = 1 - (\alpha + \beta + \gamma + \xi)$. Similar to the derivations in Chapter 2.2, the MLE estimators for $\alpha, \beta, \gamma, \xi$ are

$$\begin{aligned}
\hat{\alpha}^{MLE} &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=1\}}, & \hat{\beta}^{MLE} &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=2\}}, \\
\hat{\gamma}^{MLE} &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}}, & \hat{\xi}^{MLE} &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=4\}},
\end{aligned}$$

and $\delta_{\text{in}}, \delta_{\text{out}}$ can be obtained through solving

$$\begin{aligned}
\sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{i + \delta_{\text{in}}} - \frac{\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t \in \{3,4,5\}\}}}{\delta_{\text{in}}} - \frac{1}{n}\sum_{t=1}^{n} \frac{N(t)}{t + \delta_{\text{in}} N(t)} \mathbf{1}_{\{J_t \in \{1,2\}\}} &= 0, \\
\sum_{j=0}^{\infty} \frac{N^{\text{out}}_{>j}(n)/n}{j + \delta_{\text{out}}} - \frac{\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t \in \{1,4,5\}\}}}{\delta_{\text{out}}} - \frac{1}{n}\sum_{t=1}^{n} \frac{N(t)}{t + \delta_{\text{out}} N(t)} \mathbf{1}_{\{J_t \in \{2,3\}\}} &= 0.
\end{aligned}$$

We first naively fit the linear preferential attachment model to the full network using MLE. The MLE estimators are

$$\begin{aligned}
&(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{\rho}, \hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}}) = \\
&(3.08 \times 10^{-3}, 8.55 \times 10^{-1}, 1.39 \times 10^{-1}, 4.76 \times 10^{-5}, 3.06 \times 10^{-3}, 0.547, 0.134). \quad (2.5.1)
\end{aligned}$$

To evaluate the goodness-of-fit, 20 network realizations were simulated from the fitted model. We overlaid the empirical in- and out-degree frequencies of the original network with that of the simulations. If the model fits the data well, the degree frequencies of the data should lie within the range formed by that

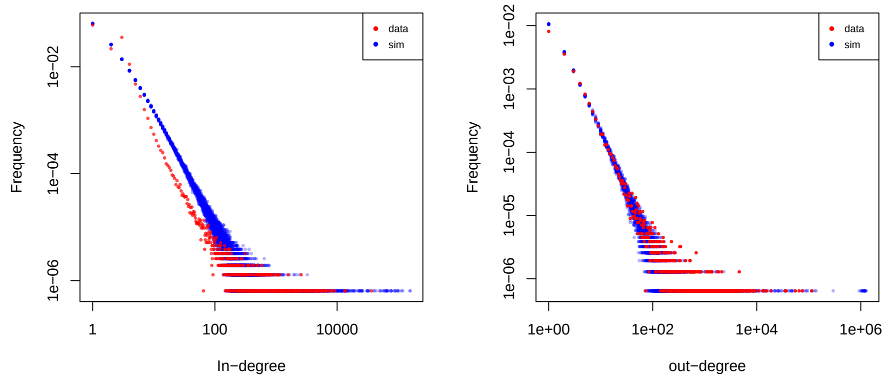


Figure 2.5.1: Empirical in- and out-degree frequencies of the full Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values (2.5.1) from MLE (blue). The scatter plots for the degree frequencies from the 20 simulations are overlaid together to form an informal confidence region for the degree distribution of the fitted model

of the simulations, which gives an informal confidence region for the degree distributions. From Figure 2.5.1, we see that while the data roughly agrees with the simulations in the out-degree frequencies, the deviation in the in-degree frequencies is noticeable.

To better understand the discrepancy in the in-degree frequencies, we examined the link data and their timestamps and discovered bursts of messages originating from certain nodes over small time intervals. According to Wikipedia policy [70], certain administrating accounts are allowed to send group messages to multiple users simultaneously. These bursts presumably represent broadcast announcements generated from these accounts. These administrative broadcasts can also be detected if we apply the linear preferential attachment model to the network in local time intervals. We divided the total time frame down to sub-intervals of varying length each containing the formation of 10^4 edges. The number 10^4 is chosen to ensure good asymptotics as shown in Table 2.4.1. This process generated 155 networks,

$$G(n_{k-1}), \ldots, G(n_k - 1), \quad k = 1, \ldots, 155.$$

For each of the 155 datasets, we fit a preferential attachment model using MLE. The resulting estimates $(\hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}})$ are plotted against the corresponding timeline on the upper left panel of Figure 2.5.2. Notice that $\hat{\delta}_{\text{in}}$ exhibits large spikes at various times. Recall from (1.1.1), a large value of δ_{in} indicates that the probability of an existing node v receiving a new message becomes less dependent on its in-degree, i.e., previous popularity. These spikes appear to be directly related to the occurrences of group messages. This plot is truncated after the day 2016/3/16, on which a massive group message of size 48,957 was sent and the model can no longer be fit.

We identified 37 users who have sent, at least once, 40 or more consecutive messages in the message history. This is evidence that group messages were sent by this user. We presume these nodes are administrative accounts; they are responsible for about 30% of the total messages sent. Since their behavior cannot be regarded as normal social interaction, we excluded messages from these accounts from the dataset in our analysis. We then also removed nodes with zero in- and out-degrees.

The re-estimated parameters after the data cleaning are displayed in the other three panels of Figure 2.5.2. Here all parameter estimates are quite stable through time. The reduced network now contains 112,919 nodes and 1,086,982 edges, to which we fit the linear preferential attachment model. The fitted pa-

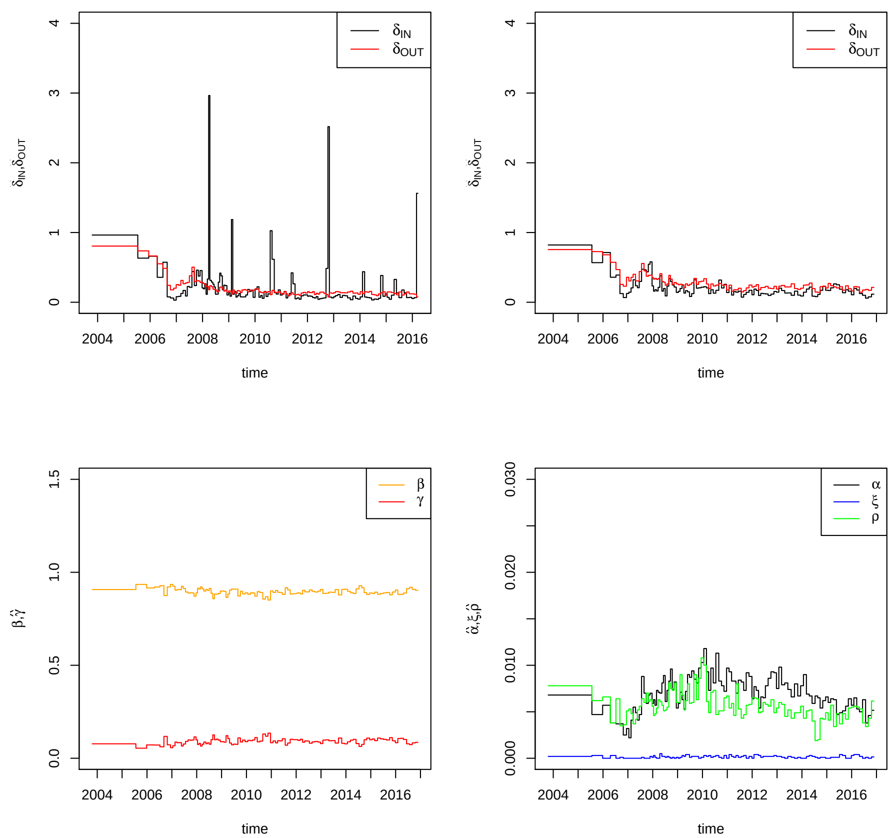


Figure 2.5.2: Local parameter estimates of the linear preferential attachment model for the full and reduced Wiki talk network. Upper left: $(\hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}})$ for the full network. Upper right, lower left, lower right: $(\hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}})$, $(\hat{\beta}, \hat{\gamma})$, $(\hat{\alpha}, \hat{\xi}, \hat{\rho})$ for the reduced network, respectively.

rameters based on MLE for our reduced dataset are

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{\rho}, \hat{\delta}_{\text{in}}, \hat{\delta}_{\text{out}}) = (6.95 \times 10^{-3}, 8.96 \times 10^{-1}, 9.10 \times 10^{-2}, 1.44 \times 10^{-4}, 5.61 \times 10^{-3}, 0.174, 0.257). \quad (2.5.2)$$

Again the degree distributions of the data and 20 simulations from the fitted model are displayed in Figure 2.5.3. The out-degree distribution of the data agrees reasonably well with the simulations. For the in-degree distribution, the fit is better than that for the entire dataset (Figure 2.5.1). However, for smaller in-

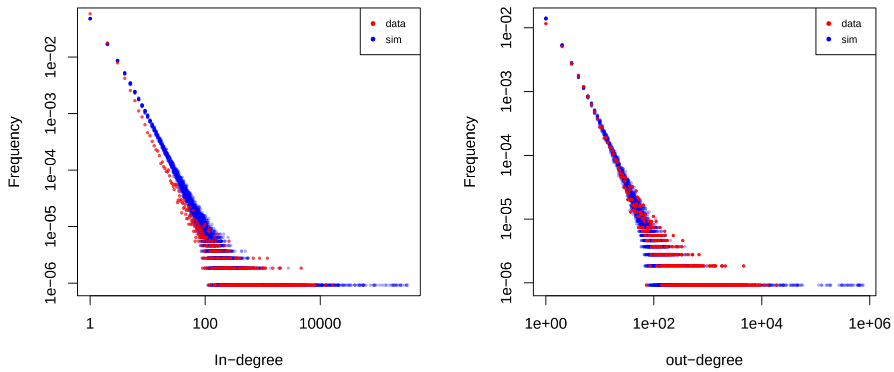


Figure 2.5.3: Empirical in- and out-degree frequencies of the reduced Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values (2.5.2) from MLE (blue).

degrees, the fitted model over-estimates the in-degree frequencies. We speculate that in many social networks, the out-degree is in line with that predicted by the preferential attachment model. An individual node would be more likely to reach out to others if having done so many times previously. For in-degrees, the situation is complicated and may depend on a multitude of factors. For instance, the choice of recipient may depend on the community that the sender is in, the topic being discussed in the message, etc. As an example a group leader might send messages to his/her team on a regular basis. Such examples violate the base assumptions of the preferential attachment model and could result in the deviation between the data and the simulations.

Next we consider the estimation method of Chapter 2.3 applied to a single snapshot of the data. In order to implement this procedure, we donned blinders and assumed that our dataset consists only of the information of the wiki data at the last timestamp. That is, information about administrative broadcasts, and

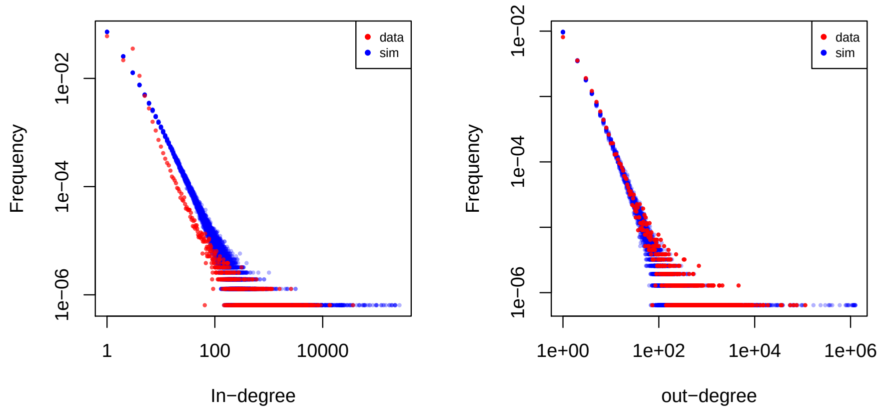


Figure 2.5.4: Empirical in- and out-degree frequencies of the full Wiki talk network (red) and that from 20 realizations of the linear preferential attachment network with fitted parameter values (2.5.3) from the snapshot estimator (blue).

other aspects of the data learned by looking at the previous history of the data are unavailable. In particular, we would have no knowledge of the existence of the two additional scenarios corresponding to $J_n = 4, 5$. With this in mind, we fit the three scenario model using the methods in Chapter 2.3. The fitted parameters are

$$(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta}_{\text{in}}, \tilde{\delta}_{\text{out}}) = (5.80 \times 10^{-4}, 8.55 \times 10^{-1}, 1.45 \times 10^{-1}, 0.199, 0.165). \qquad (2.5.3)$$

The comparison of the degree distributions between the data and simulations from the fitted model is displayed in Figure 2.5.4 and is not too dissimilar to the plots in Figure 2.5.1 that are based on maximum likelihood estimation using the full network data. In particular, the out-degree distribution is matched reasonably well, but the fitted model does a poor job of capturing the in-degree distribution.

We see from this example that while the linear preferential attachment model is perhaps too simplistic for the Wiki talk network dataset, it has the ability to illuminate some gross features, such as the out-degrees, as well as to capture important structural changes such as the group message behavior. Consequently, despite its limitation, this model may be used as a building block for more flexible models. Modification to the existing model formulation and more careful analysis of change points in parameters is a direction for future research.

2.6 Estimation Using Extreme Value Theory

Often it is difficult to believe in the existence of a true model, especially one whose parameters remain constant over time. Allowing, as we do, a preferential attachment model with only a few parameters and no possibility for node removal may seem simplistic and unrealistic for social network data. Of course, preferential attachment is only one mechanism for network formation and evidence for its use in fields outside data networks is mixed [31, 32] and we restrict attention to linear preferential attachment. Even imperfect models have the potential to capture salient properties in the data, such as heavy-tailedness of the in-degree and out-degree distributions, and to identify departures from model assumptions.

While maximum likelihood estimation is essentially the gold standard for cases when the underlying model is a good representation of the data, it may perform poorly in case the model is far from being appropriate. In another study [63], we consider a semi-parametric estimation approach for network models that exhibit heavy-tailed degree distributions. This alternative estimation methodology borrows ideas from extreme value theory. From now on, we use MLE, SN and EV to denote the MLE, one-snapshot and extreme value estimation methods, respectively.

For a given graph, the data we utilize to do estimation are the in- and out-degrees of each node in the snapshot at a fixed time point. In Chapter 2.6.1, we describe the Hill estimation for the marginal tail indices of the in- and out-degrees. In Chapters 2.6.2 and 2.6.3, we capitalize on the multivariate regular variation asymptotic dependence structure of in- and out-degree.

Given a graph $G(n)$ at time n, the in- and out-degrees for each node are denoted by $(D_v^{\text{in}}(n), D_v^{\text{out}}(n))$, $v = 1, \ldots, N(n)$. Let $F_n(\cdot)$ be the empirical distribution of this data on $\mathbb{N} \times \mathbb{N}$. Then from (1.1.3), almost surely F_n converges weakly to a limit distribution F on $\mathbb{N} \times \mathbb{N}$ which is the measure corresponding to the mass function $\{p_{ij}\}$. Let $\epsilon_{(i,j)}(\cdot)$ be the Dirac measure concentrating on (i, j) and we have from (1.1.3), with probability 1,

$$F_n(\cdot) = \frac{1}{N(n)} \sum_{v=1}^{N(n)} \epsilon_{(D_v^{\text{in}}(n), D_v^{\text{out}}(n))}(\cdot) = \sum_{i,j} \frac{N_n(i, j)}{N(n)} \epsilon_{(i,j)}(\cdot) \xrightarrow{w} \sum_{i,j} p_{ij} \epsilon_{(i,j)}(\cdot) =: F(\cdot). \quad (2.6.1)$$

Convergence results for the joint tail empirical distribution when $\{D_v^{\text{in}}(n), D_v^{\text{out}}(n), 1 \leq v \leq N(n)\}$ are scaled by growing functions of n are given in [65].

2.6.1 Estimating tail indices; Hill estimation.

We have seen that $F_{\text{in}}(\cdot) := \sum_i p_i^{\text{in}} \epsilon_i(\cdot)$ has a regularly varying tail with index $-\iota_{\text{in}}$. One common way to estimate tail indices such as ι_{in} is to use the Hill estimator [15, 28, 50], but for non-iid network data, use of Hill's estimator requires justification. For the models being considered, this issue is resolved in [65, 68], where we prove consistency of the Hill estimator for data generated from linear PA models. Hence, we proceed to estimate ι_{in} and ι_{out} by the corresponding Hill estimator. For ι_{in} proceed as follows. Let $D^{\text{in}}_{(1)}(n) \geq \ldots \geq D^{\text{in}}_{(N(n))}(n)$ be the decreasing order statistics of $D^{\text{in}}_v(n)$, $v = 1, \ldots, N(n)$. The Hill estimator $\hat{\iota}_{\text{in}}(k_n)$ based on k_n largest degrees is

$$\hat{\iota}_{\text{in}}(k_n) = \left(\frac{1}{k_n} \sum_{j=1}^{k_n} \left(\log(D^{\text{in}}_{(j)}(n)) - \log(D^{\text{in}}_{(k_n+1)}(n)) \right) \right)^{-1}, \tag{2.6.2}$$

where $\{k_n\}$ is an intermediate sequence satisfying $k_n \to \infty$ and $k_n/n \to 0$. The estimate of ι_{out} is defined similarly. To select k_n in practice, [10] proposed computing the Kolmogorov-Smirnov (KS) distance between the empirical distribution tail of the upper k observations and the power-law distribution with index $\hat{\iota}_{\text{in}}(k)$:

$$D_k := \sup_{y \geq 1} \left| \frac{1}{k} \sum_{j=1}^{k} \mathbf{1}_{\{D^{\text{in}}_{(j)}(n)/D^{\text{in}}_{(k+1)}(n) > y\}} - y^{-\hat{\iota}_{\text{in}}(k)} \right|, \quad 1 \leq k \leq n-1.$$

Then the optimal k^* is the one that minimizes the KS distance

$$k^* := \operatorname*{argmin}_{1 \leq k \leq n} D_k,$$

and the tail index is estimated by $\hat{\iota}_{\text{in}}(k^*)$. We refer to the above procedure as the *minimum distance method*. It is widely used by data repositories of large network datasets such as KONECT (`http://konect.cc/`) [42] and is realized in the R-package *poweRlaw* [24]. In [17], we find experimentally that the minimum distance method is effective for simulated data from the linear PA model. Furthermore, degree counts are discrete and do not exactly comply with the Pareto assumption made in [10], but experiments [17] show that the method of combining Hill estimation with the minimum distance technique works well in practice; see Chapter 2.7. Theoretical justifications of this procedure for iid data will be given in Chapter 4.

2.6.2 Estimating dependency between in- and out-degrees

Since the limiting random vector $(I, O) \sim F$ corresponding to p_{ij} in (1.1.3) is jointly regularly varying and satisfies

$$t\mathbf{P}\left[\left(\frac{I^a}{t^{1/\iota_{\text{out}}}}, \frac{O}{t^{1/\iota_{\text{out}}}}\right) \in \cdot\right] \to \tilde{\nu}(\cdot), \quad \text{in } \mathbb{M}(\mathbb{R}_+^2 \setminus \{\mathbf{0}\}), \tag{2.6.3}$$

we apply a polar coordinate transformation, for example, with the L_2-norm,

$$(I^a, O) \mapsto (\sqrt{I^{2a} + O^2}, \arctan(O/I^a)) := (R, T),$$

where $a = \iota_{\text{in}}/\iota_{\text{out}}$. Then, with respect to F in (2.6.1), the conditional distribution of T given $R > r$ converges weakly (see, for example, [50, p. 173]),

$$F[T \in \cdot | R > r] \to S(\cdot), \quad r \to \infty,$$

where S is the *angular measure* and describes the asymptotic dependence of the standardized pair (I^a, O). Since for large r, $F[T \in \cdot | R > r] \approx S(\cdot)$ and for large n, $F_n \approx F$, it is plausible that for r and n large $F_n[T \in \cdot | R > r] \approx S(\cdot)$. Refer to [50, p. 307] for a more precise argument and recall F_n is the empirical measure defined in (2.6.1).

Based on observed degrees $\{(D_v^{\text{in}}(n), D_v^{\text{out}}(n)); v = 1, \dots, N(n)\}$, how does this work in practice? First a is replaced by $\hat{a} = \hat{\iota}_{\text{in}}/\hat{\iota}_{\text{out}}$ estimated from Chapter 2.6.1. Then the distribution S is estimated via the empirical distribution of the sample angles $T_n(v) := \arctan(D_v^{\text{out}}(n)/D_v^{\text{in}}(n)^{\hat{a}})$ for which $R_v(n) := \sqrt{D_v^{\text{in}}(n)^{2\hat{a}} + D_v^{\text{out}}(n)^2} > r$ exceeds some large threshold r. This is the Peaks Over Threshold (POT) methodology commonly employed in extreme value theory [11].

For the linear PA model, the form of S is known and there is a density that is an explicit function of model parameters [55]. After estimating ι_{in} and ι_{out} by the minimum distance method, the remaining parameters can then be estimated by an approximate likelihood method that we now explain.

2.6.3 Extreme value estimation for the linear PA model

From (1.1.5) and (1.1.6),

$$\delta_{\text{in}} = \frac{\iota_{\text{in}}(\alpha+\beta)-1}{\alpha+\gamma}, \quad \delta_{\text{out}} = \frac{\iota_{\text{out}}(\beta+\gamma)-1}{\alpha+\gamma},$$

so that the linear PA model may be parameterized by $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \iota_{\text{in}}, \iota_{\text{out}})$. To construct the EV estimates, begin by computing the minimum distance estimates $\hat{\iota}_{\text{in}}^{EV}, \hat{\iota}_{\text{out}}^{EV}$ of the in- and out-degree indices. The parameter β, which represents the proportion of edges connected between existing nodes, is estimated by $\hat{\beta}^{EV} = 1 - N(n)/n$.

From (2.6.3), $\arctan(O/I^a)$ given $I^{2a} + O^2 > r^2$ converges weakly as $r \to \infty$ to the distribution of a random variable Θ [55, Section 4.1.2], whose pdf is given by ($0 \le x \le \pi/2$)

$$\begin{aligned} f_\Theta(x; \alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}}) \quad \propto \quad & \frac{\gamma}{\delta_{\text{in}}}(\cos x)^{\frac{\delta_{\text{in}}+1}{a}-1}(\sin x)^{\delta_{\text{out}}-1}\int_0^\infty t^{\iota_{\text{in}}+\delta_{\text{in}}+a\delta_{\text{out}}}e^{-t(\cos x)^{1/a}-t^a \sin x}\mathrm{d}t \\ & + \frac{\alpha}{\delta_{\text{out}}}(\cos x)^{\frac{\delta_{\text{in}}}{a}-1}(\sin x)^{\delta_{\text{out}}}\int_0^\infty t^{a-1+\iota_{\text{in}}+\delta_{\text{in}}+a\delta_{\text{out}}}e^{-t(\cos x)^{1/a}-t^a \sin x}\mathrm{d}t. \end{aligned} \tag{2.6.4}$$

By replacing $\beta, \iota_{\text{in}}, \iota_{\text{out}}$ with their estimated values $\hat{\beta}^{EV}$, $\hat{\iota}_{\text{in}}^{EV}$, and $\hat{\iota}_{\text{out}}^{EV}$ and setting $\gamma = 1 - \alpha - \hat{\beta}^{EV}$, the density (2.6.4) can be viewed as a profile likelihood function (based on a single observation x) of the unknown parameter α, which we denote by

$$l(\alpha; x) = f_\Theta(x; \alpha, \hat{\beta}^{EV}, 1 - \alpha - \hat{\beta}^{EV}, \hat{\delta}_{\text{in}}^{EV}, \hat{\delta}_{\text{out}}^{EV}).$$

Given the degrees $((D_v^{\text{in}}(n), D_v^{\text{out}}(n)), v \in V(n))$, $\hat{\alpha}^{EV}$ can be computed by maximizing the profile likelihood based on the observations $(D_v^{\text{in}}(n), D_v^{\text{out}}(n))$ for which $R_v(n) > r$ for a large threshold r. That is,

$$\hat{\alpha}^{EV} := \underset{0\leq\alpha\leq 1}{\operatorname{argmax}} \sum_{v=1}^{N(n)} \log l\left(\alpha; \arctan\left(\frac{D_v^{\text{out}}(n)}{(D_v^{\text{in}}(n))^{\hat{a}}}\right)\right) \mathbf{1}_{\{R_v(n)>r\}}, \tag{2.6.5}$$

where r is typically chosen as the $(n_{\text{tail}} + 1)$-th largest $R_v(n)$'s for a suitable n_{tail}. This estimation procedure is sometimes referred to as the "independence estimating equations" (IEEs) method [8, 61], in which the dependence between observations is ignored. This technique is often used when the joint distribution of the data is unknown or intractable. Finally, using the constraint, $\alpha + \beta + \gamma = 1$, we estimate γ by $\hat{\gamma}^{EV} = 1 - \hat{\alpha}^{EV} - \hat{\beta}^{EV}$.

2.7 Extreme value estimation results

In this section, we demonstrate the estimation of the linear PA and related models through the EV method described in Chapter 2.6.3. In Chapter 2.7.1, data are simulated from the standard linear PA model and used to estimate the true parameters of the underlying model. Chapter 2.7.2 considers data generated from the linear PA model but corrupted by random addition or deletion of edges. Our goal is to estimate the parameters of the original linear PA model.

Throughout the section, the EV method is compared with the two parametric approaches, the MLE and SN methods, under the linear PA model. Note that a main difference between the MLE, SN and EV methods lies in the amount of data utilized. The MLE approach requires the entire growth history of the network while the SN method uses only a single snapshot of the network. The EV method starts with thresholded data to estimate tail indices and then proceeds to estimate two more parameters of the re-parametrized model. When the data generating model is correct, MLE is certainly the most efficient, but requires a complete historical data set. As we shall see, in the case where the data is corrupted, the EV method provides an attractive and reliable alternative.

2.7.1 Estimation for the linear PA model

Comparison of EV with MLE and SN

Figure 2.7.1 presents biases for estimates of $(\alpha, \iota_{\text{in}}, \iota_{\text{out}})$ using EV, MLE, and SN methods on data simulated from the linear PA model.

We held $(\beta, \delta_{\text{in}}, \delta_{\text{out}}) = (0.4, 1, 1)$ constant and varied $\alpha = 0.1, 0.2, 0.3, 0.4$ so that the true values of $\gamma, \iota_{\text{in}}, \iota_{\text{out}}$ were also varying. For each set of parameter values $(\alpha, \iota_{\text{in}}, \iota_{\text{out}})$, 200 independent replications of a linear PA network with $n = 10^5$ edges were simulated and the true values of $(\iota_{\text{in}}, \iota_{\text{out}})$ were computed by (1.1.5) and (1.1.6). We estimated $(\iota_{\text{in}}, \iota_{\text{out}})$ by the minimum distance method $(\hat{\iota}_{\text{in}}^{EV}, \hat{\iota}_{\text{out}}^{EV})$, MLE and the one-snapshot methods applied to the parametric model (cf. Chapters 2.2 and 2.3), denoted by $(\hat{\iota}_{\text{in}}^{MLE}, \hat{\iota}_{\text{out}}^{MLE})$ and $(\hat{\iota}_{\text{in}}^{SN}, \hat{\iota}_{\text{out}}^{SN})$, respectively. With $(\hat{\iota}_{\text{in}}^{EV}, \hat{\iota}_{\text{out}}^{EV})$, $\hat{\alpha}^{EV}$ is calculated by (2.6.5) using $n_{\text{tail}} = 200$.

As seen here, for simulated data from a known model, MLE outperforms other estimation procedures. The EV procedure tends to have much larger variance than both MLE and SN with slightly more bias. This is not surprising as the performance of the EV estimators is dependent on the quality of the following approximations:

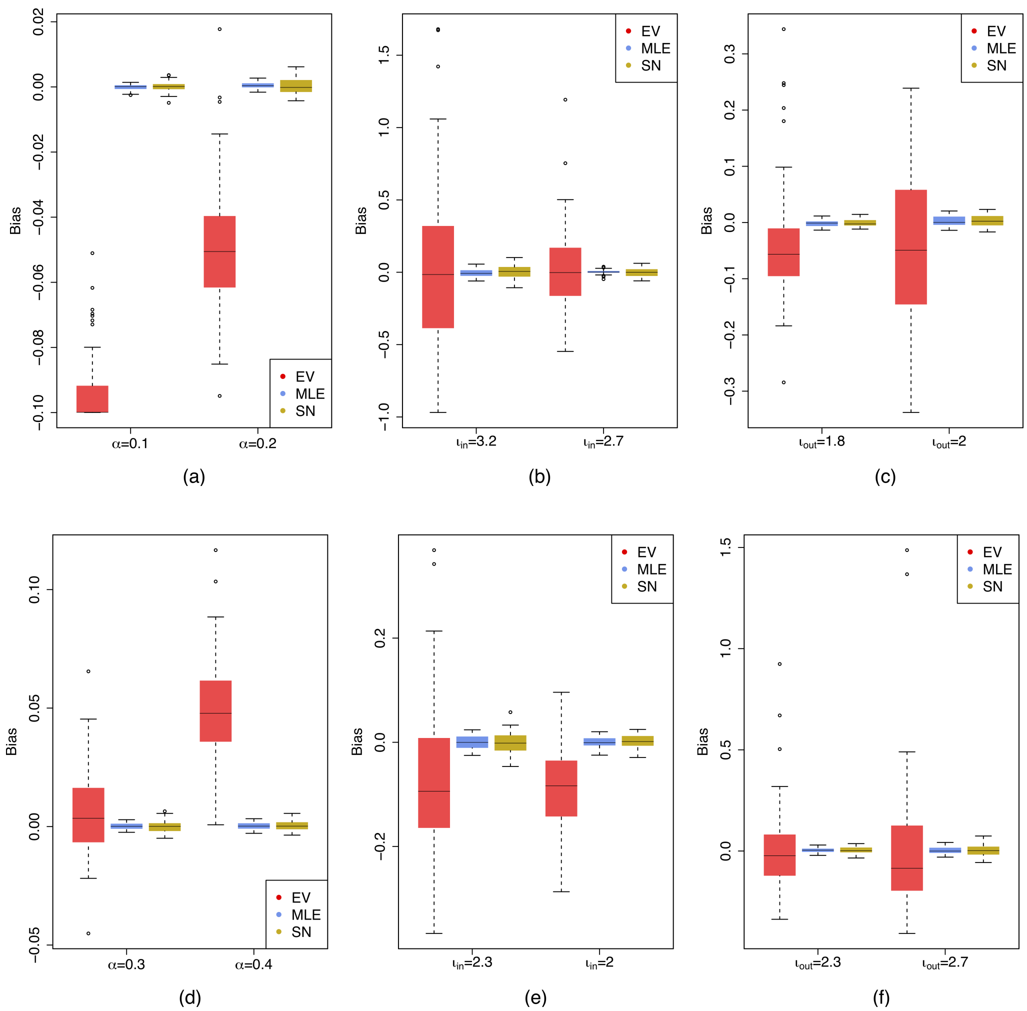


Figure 2.7.1: Boxplots of biases for estimates of $(\alpha, \iota_{\text{in}}, \iota_{\text{out}})$ using EV, MLE and SN methods. Panels (a)–(c) correspond to the case where $\alpha = 0.1, 0.2$ and (d)–(f) are for $\alpha = 0.3, 0.4$, holding $(\beta, \delta_{\text{in}}, \delta_{\text{out}}) = (0.4, 1, 1)$ constant.

1. The number of edges in the network, n, should be sufficiently large to ensure a close approximation of $N_n(i, j)/N(n)$ to the limit joint pmf p_{ij}.

2. The choice of thresholds must guarantee the quality of the EV estimates for the indices and the limiting angular distribution. The thresholding means some estimates are based on only a small fraction of the data and hence have large uncertainty.

3. The parameter a used to transform the in- and out-degrees to standard regular variation is estimated and thus subject to estimation error which propagates throughout the remaining estimation procedures.

2.7.2 Data corrupted by random edge addition/deletion.

PA models are designed to describe human interaction in social networks but what if data collected from a network is corrupted or usual behavior is changed? Corruption could be due to collection error and atypical behavior could result from users hiding their network presence or trolls acting as provocateurs. In such circumstances, the task is to unmask data corruption or atypical behavior and recover the parameters associated with the original preferential attachment rules.

In the following, we consider network data that are generated from the linear PA model but corrupted by random addition or deletion of edges. For such corrupted data, we attempt to recover the original model and compare the performances of MLE, SN, and EV methods.

Randomly adding edges.

We consider a network generating algorithm with linear PA rules but also a possibility of adding random edges. Let $G(n) = (V(n), E(n))$ denote the graph at time n. We assume that the edge set $E(n)$ can be decomposed into two disjoint subsets: $E(n) = E^{PA}(n) \bigcup E^{RA}(n)$, where $E^{PA}(n)$ is the set of edges resulting from PA rules, and $E^{RA}(n)$ is the set of those resulting from random attachments. This can be viewed as an interpolation of the PA network and the Erdös-Rényi random graph.

More specifically, consider the following network growth. Given $G(n-1)$, $G(n)$ is formed by creating a new edge where:

(1) With probability p_a, two nodes are chosen randomly (allowing repetition) from $V(n-1)$ and an edge is created connecting them. The possibility of a self loop is allowed.

(2) With probability $1-p_a$, a new edge is created according to the preferential attachment scheme $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}})$ on $G^{PA}(n-1) := (V(n-1), E^{PA}(n-1))$.

The question of interest is, if we are unaware of the perturbation effect and pretend the data from this model are coming from the linear PA model, can we recover the PA parameters? To investigate, we generate networks of $n = 10^5$ edges with parameter values

$$(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}}) = (0.3, 0.4, 0.3, 1, 1), \quad p_a \in \{0.025, 0.05, 0.075, 0.1, 0.125, 0.15\}.$$

For each network, the original PA model is fitted using the MLE, SN and EV methods, respectively. The angular MLE in (2.6.5) in the extreme value estimation is performed based on $n_{\text{tail}} = 500$ tail observations. In order to compare these estimators, we repeat the experiment 200 times for each value of p_a and obtain 200 sets of estimated parameters for each method. Figure 2.7.2 summarizes the estimated values for $(\delta_{\text{in}}, \delta_{\text{out}}, \alpha, \gamma, \iota_{\text{in}}, \iota_{\text{out}})$ for different values of p_a. The mean estimates are marked by crosses and the 2.5% and 97.5% empirical quantiles are marked by the bars. The true value of parameters are shown as the horizontal lines.

While all parameters deviate from the true value as p_a increases and the network becomes more "noisy", the EV estimates for $(\delta_{\text{in}}, \delta_{\text{out}})$ exhibit smaller bias than the MLE and SN methods (Figure 2.7.2 (a) and (b)). All three methods give underestimated probabilities (α, γ) (Figure 2.7.2 (c) and (d)). This is because the perturbation step (1) creates more edges between existing nodes and consequently inflates the estimated value of β.

Also note that the mean EV estimates of $(\iota_{\text{in}}, \iota_{\text{out}})$ stay close to the theoretical values for all choices of p_a; see Figure 2.7.2 (e) and (f). The MLE and SN estimates of $(\iota_{\text{in}}, \iota_{\text{out}})$, which are computed from the corresponding estimates for $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}})$, show strong bias as p_a increases. In this case, the EV method is robust for estimating the PA parameters and recovering the tail indices from the original model.

Randomly deleting edges.

We now consider the scenario where a network is generated from the linear PA model, but a random proportion p_d of edges are deleted at the final time. We do this by generating $G(n)$ and then deleting $[np_d]$ edges by sampling for $E(n)$ without replacement. For the simulation, we generated networks with parameter values

$$(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}}) = (0.3, 0.4, 0.3, 1, 1), \quad p_d \in \{0.025, 0.05, 0.075, 0.1, 0.125, 0.15\}.$$

Again, for each p_d, the experiment is repeated 200 times and the resulting parameter plots are shown in Figure 2.7.3 using the same format as for Figure 2.7.2. For the EV method, 100 tail observations were used to compute an $\hat{\alpha}^{EV}$.

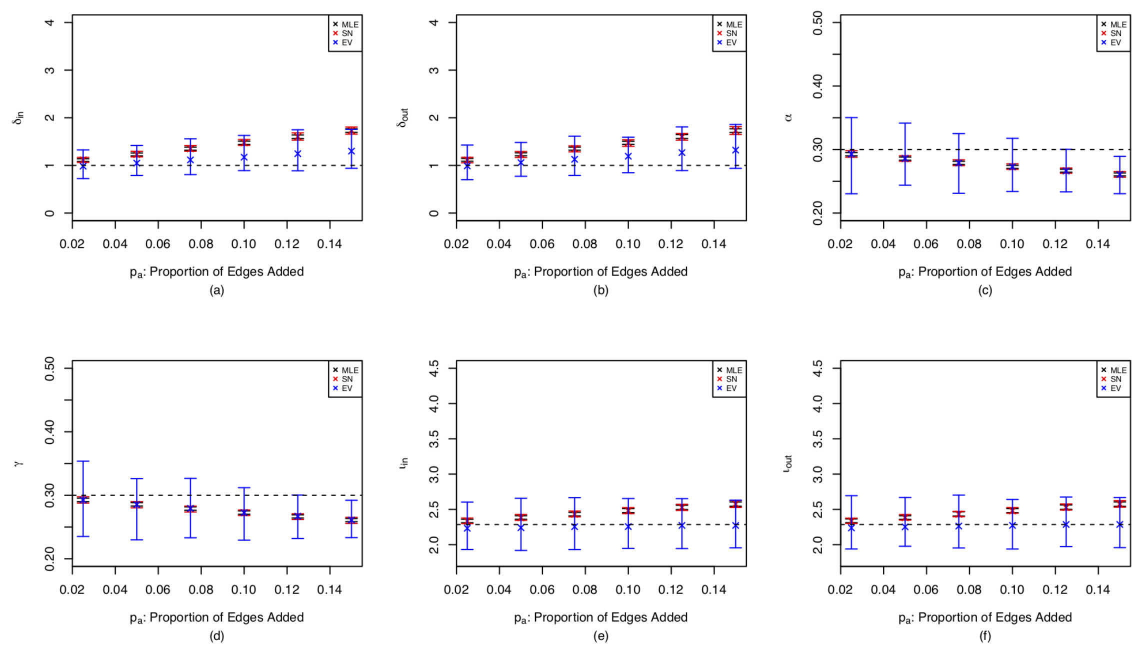


Figure 2.7.2: Mean estimates and 2.5% and 97.5% empirical quantiles of (a) δ_{in}; (b) δ_{out}; (c) α; (d) γ; (e) ι_{in}; (f) ι_{out}, using MLE (black), SN (red) and EV (blue) methods over 200 replications, where $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}}) = (0.3, 0.4, 0.3, 1, 1)$ and $p_a = 0.025, 0.05, 0.075, 0.1, 0.125, 0.15$. For the EV method, 500 tail observations were used to obtain $\hat{\alpha}^{EV}$.

Surprisingly, for all six parameters considered, MLE estimates stay almost unchanged for different values of p_d while SN and EV estimates underestimate $(\delta_{\text{in}}, \delta_{\text{out}})$ and overestimate (α, γ), with increasing magnitudes of biases as p_d increases. For tail estimates, the minimum distance method still gives reasonable results (though with larger variances), whereas the SN method keeps underestimating ι_{in} and ι_{out}.

The performance of MLE in this case is surprisingly competitive. This is intriguing and in ongoing work, we will think about why this is the case.

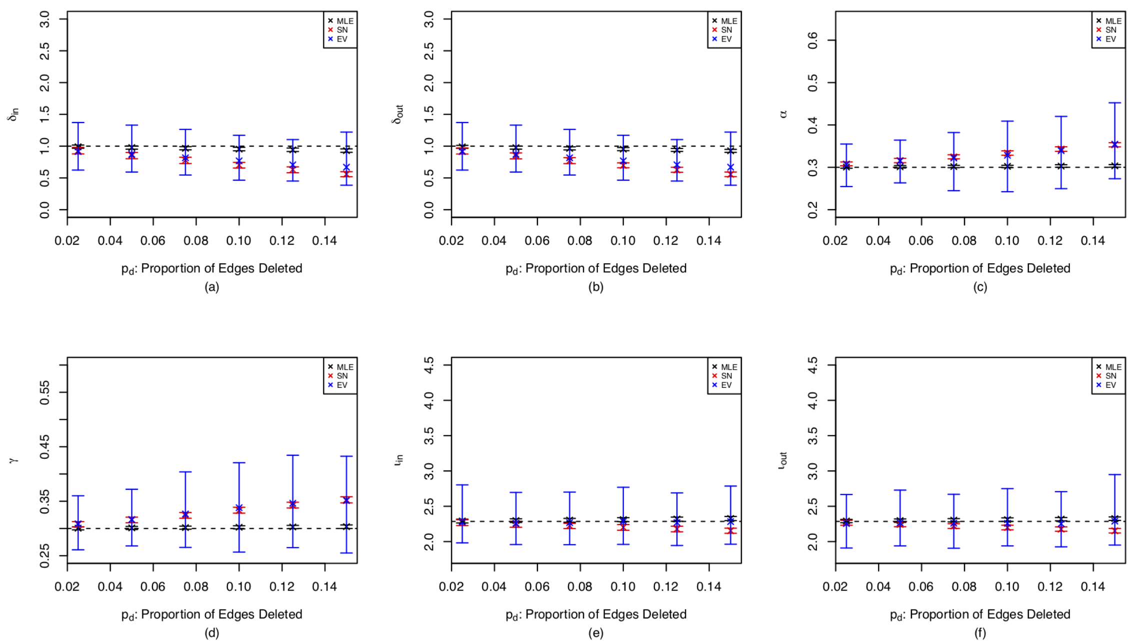


Figure 2.7.3: Mean estimates and 2.5% and 97.5% empirical quantiles of (a) δ_{in}; (b) δ_{out}; (c) α; (d) γ; (e) ι_{in}; (f) ι_{out}, using MLE (black), SN (red) and EV (blue) methods over 50 replications, where $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}}) = (0.3, 0.4, 0.3, 1, 1)$ and $p_d = 0.025, 0.05, 0.075, 0.1, 0.125, 0.15$. For the EV method, 100 tail observations were used to compute $\hat{\alpha}^{EV}$.

CHAPTER 3

DEGREE GROWTH RATES AND INDEX ESTIMATION IN A DIRECTED LINEAR PA MODEL

3.1 Overview

Empirical studies on social network data often reveal that in- and out-degree distributions marginally follow power laws. Theoretically, this is also true for *linear* preferential attachment models, which makes preferential attachment appealing in network modeling; see [7, 39, 40] for references. Also, the empirical joint degree frequency converges to the probability mass function (pmf) of a pair of limit random variables that are jointly regularly varying (cf. [40, 52, 55, 67]). However, questions related to joint degree growth and index estimation still remain unresolved. In this chapter, we focus on three main problems:

1. For a fixed node in a linear preferential attachment graph, what is the joint behavior of in- and out-degree as the graph size grows?
2. What are the convergence properties of the tail empirical joint measure of in- and out-degrees indexed by node?
3. When estimating the marginal power-law indices of in- and out-degree, can we use the Hill estimator as a consistent estimator?

What is the justification for interest in Hill estimation of power-law indices for network data? Repositories of large network datasets such as KONECT (`http://konect.cc/`, [42]) provide summary statistics for all the archived network datasets and among the summary statistics are estimates of degree indices computed with Hill estimators, despite the fact that evidence for Hill estimator consistency is scant for network data [68].

Another justification is robust parameter estimation methods in network models based on extreme value techniques. In [63], we couple the Hill estimation of marginal degree distribution tail indices with a minimum distance threshold selection method introduced in [10] and compare this method with the parametric estimation approaches used in [64]. The Hill estimation is more robust against modeling error and data corruption. Therefore, an affirmative answer to the third question helps justify all of these inference methodologies.

In the directed case, consistency of the two marginal Hill estimators results from resolving the first two questions, since in a similar vein to [68], we consider the Hill estimator as a functional of the marginal tail empirical measure. So convergence results of marginal tail empirical measures lead to the consistency of Hill estimators by a mapping argument.

To answer the first question about degree behavior of fixed nodes as graph size grows, we mimic in- and out-degree growth of a fixed node using pairs of *switched birth processes with immigration* (SBI processes). The SBI processes use Bernoulli switching between pairs of independent *birth processes with immigration* (BI processes). We embed the directed network growth model into a sequence of paired SBI processes. Whenever a new node is added to the network, a new pair of SBI processes is initiated. Using convergence results for BI processes (cf. [49, Chapter 5.11], [57, 68]), we give the joint limits of the in- and out-degrees of a fixed node as well as the joint maximal degree growth. Proving the convergence of the tail empirical joint measure in the second question requires showing concentration results for degree counts compared with expected degree counts. With embedding techniques, we prove the limit distribution of the empirical joint degree frequencies in a way that is different from the one used in [55], and then justify the concentration results.

3.1.1 Background

Our approach to the Hill estimator considers it as a functional of the tail empirical measure so we start with necessary background and review standard results (cf. [50, Chapter 3.3.5 and 6.1.4]).

Non-standard regular variation

Let $M_+([0,\infty]^2 \setminus \{\mathbf{0}\})$ be the set of Radon measures on $[0,\infty]_+^2 \setminus \{\mathbf{0}\}$. Then a random vector (X, Y) is *non-standard regularly varying* on $[0,\infty]_+^2 \setminus \{\mathbf{0}\}$ if there exist scaling functions $b_i(t) \to \infty$, $i = 1, 2$ such that as $t \to \infty$,

$$t\mathbf{P}\left[\left(\frac{X}{b_1(t)}, \frac{Y}{b_2(t)}\right) \in \cdot\right] \xrightarrow{v} \nu(\cdot), \quad \text{in } M_+([0,\infty]^2 \setminus \{\mathbf{0}\}), \tag{3.1.1}$$

where $\nu(\cdot) \in M_+([0,\infty]^2 \setminus \{\mathbf{0}\})$ is called the limit or tail measure [52, 55], and "$\xrightarrow{v}$" denotes the vague convergence of measures in $M_+([0,\infty]^2 \setminus \{\mathbf{0}\})$. The phrasing in (3.1.1) implies the marginal distributions have regularly varying tails.

Hill Estimator

For $x \in (0,\infty]$, define the measure $\epsilon_x(\cdot)$ on Borel subsets A of $(0,\infty]$ by

$$\epsilon_x(A) = \begin{cases} 1 & x \in A, \\ 0 & x \notin A, \end{cases} \qquad \text{for } A \in \mathcal{E}.$$

Let $M_+((0,\infty])$ be the set of non-negative Radon measures on $(0,\infty]$. A point measure m is an element of $M_+((0,\infty])$ of the form

$$m = \sum_i \epsilon_{x_i}. \tag{3.1.2}$$

For $\{X_n, n \geq 1\}$ iid and non-negative with common regularly varying distribution tail $\overline{F} \in RV_{-\iota}$, $\iota > 0$, there exists a sequence $\{b(n)\}$ satisfying $P[X_1 > b(n)] \sim 1/n$, such that for any $k_n \to \infty$, $k_n/n \to 0$,

$$\frac{1}{k_n} \sum_{i=1}^{n} \epsilon_{X_i/b(n/k_n)} \Rightarrow \nu_\iota, \qquad \text{in } M_+((0,\infty]), \tag{3.1.3}$$

where the limit measure ν_ι satisfies $\nu_\iota(y,\infty] = y^{-\iota}$, $y > 0$.

Define the Hill estimator $H_{k,n}$ based on k upper order statistics of $\{X_1, \ldots, X_n\}$ as [28]

$$H_{k,n} := \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}}, \tag{3.1.4}$$

where $X_{(1)} \geq X_{(2)} \geq \ldots \geq X_{(n)}$ are order statistics of $\{X_i : 1 \leq i \leq n\}$. In the iid case there are many proofs of consistency [12, 16, 26, 45, 46]: For $k = k_n \to \infty$, $k_n/n \to 0$, we have

$$H_{k_n,n} \xrightarrow{p} 1/\iota \qquad \text{as } n \to \infty. \tag{3.1.5}$$

The treatment in [50, Theorem 4.2] approaches consistency by showing (3.1.5) follows from (3.1.3) and we follow this approach for the network context where the iid case is inapplicable.

Node degrees

The next section constructs a directed preferential attachment model, and gives behavior of $(D_v^{\text{in}}(n), D_v^{\text{out}}(n))$, the in- and out-degrees of node v at the nth stage of construction. These degrees when scaled by appropriate powers of n (see (3.3.12)) have limits and Theorem 3.4.4 shows that the degree sequences $(D_v^{\text{in}}(n), D_v^{\text{out}}(n))_{1\le v\le n}$ have a joint tail empirical measure

$$\frac{1}{k_n}\sum_v \epsilon_{\left(D_v^{\text{in}}(n)/b_1(n/k_n),D_v^{\text{out}}(n)/b_2(n/k_n)\right)} \tag{3.1.6}$$

that converges weakly to some limit measure in $M_+([0,\infty]^2 \setminus \{\mathbf{0}\})$, where $b_1(n), b_2(n)$ are appropriate power law scaling functions and k_n is some intermediate sequence such that

$$k_n/n \to 0, \qquad k_n \to \infty, \quad \text{as } n \to \infty.$$

It also follows from (3.1.6) that for some tail indices ι_{in}, ι_{out}, and intermediate sequence k_n,

$$\frac{1}{k_n}\sum_v \epsilon_{D_v^{\text{in}}(n)/b_1(n/k_n)} \Rightarrow \nu_{\iota_{\text{in}}}, \quad \text{in } M_+((0,\infty]), \tag{3.1.7}$$

$$\frac{1}{k_n}\sum_v \epsilon_{D_v^{\text{out}}(n)/b_2(n/k_n)} \Rightarrow \nu_{\iota_{\text{out}}}, \quad \text{in } M_+((0,\infty]). \tag{3.1.8}$$

This leads to consistency of the Hill estimator for ι_{in} and ι_{out}.

3.2 Preliminaries

3.2.1 Model Construction

In this section, we consider a special case of the linear preferential attachment model by setting $\beta = 0$. One way to formally construct the model which helps with proofs is by using independent exponential random variables (r.v.'s). Define derived parameters

$$c_{\text{in}} = \frac{\alpha}{1+\delta_{\text{in}}} \qquad \text{and} \qquad c_{\text{out}} = \frac{\gamma}{1+\delta_{\text{out}}}, \tag{3.2.1}$$

and for $n \geq 1$, we will recursively define what corresponds to the in- and out-degree sequences as random elements of $(\mathbb{N}_+^2)^\infty$,

$$\mathcal{D}(n) := ((D_1^{\text{in}}(n), D_1^{\text{out}}(n)), \ldots, (D_n^{\text{in}}(n), D_n^{\text{out}}(n)), (0,0), \ldots) \tag{3.2.2}$$

with initialization

$$\mathcal{D}(1) = ((1,1), (0,0), \ldots) \tag{3.2.3}$$

corresponding to assuming $G(0)$ has a single node with a self loop. For $k \geq 1$, the recursive definition of $\{\mathcal{D}(n)\}$ uses the variables

$$\boldsymbol{e}_k^{\text{in}} := ((0,0), \ldots, (0,0), \underbrace{(1,0)}_{k\text{-th entry}}, (0,0), \ldots), \tag{3.2.4}$$

$$\boldsymbol{e}_k^{\text{out}} := ((0,0), \ldots, (0,0), \underbrace{(0,1)}_{k\text{-th entry}}, (0,0), \ldots), \tag{3.2.5}$$

and relies on competitions from exponential alarm clocks based on $\{E_k^{(n)} : k \geq 1, n \geq 1\}$, a sequence of iid standard exponential r.v.'s. Assuming $\mathcal{D}(n)$ has been given, $\mathcal{D}(n+1)$ requires $\mathcal{D}(n)$ and the $2n$ variables $\{E_j^{(n)}, j = 1, \ldots, 2n\}$ which are independent of $\mathcal{D}(n)$ and we define

$$\begin{aligned}
\overline{E}_k^{(n)} &:= \frac{E_k^{(n)}}{\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(D_k^{\text{in}}(n) + \delta_{\text{in}})}, \qquad k = 1, \ldots, n, \\
\overline{E}_k^{(n)} &:= \frac{E_k^{(n)}}{\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(D_k^{\text{out}}(n) + \delta_{\text{out}})}, \qquad k = n+1, \ldots, 2n.
\end{aligned}$$

Conditionally on $\mathcal{D}(n)$, use the $\{\overline{E}_k^{(n)} : k = 1, \ldots, 2n\}$ to create a competition between exponentially distributed alarm clocks. For $\delta_{\text{in}}, \delta_{\text{out}} > 0$ and $n \geq 1$, define choice variables

$$L_{n+1} = \sum_{l=1}^{n} l \mathbf{1}_{\left\{\overline{E}_l^{(n)} < \bigwedge_{k=1,k\neq l}^{2n} \overline{E}_k^{(n)},\, 1 \leq l \leq n\right\}} + \sum_{l=n+1}^{2n} l \mathbf{1}_{\left\{\overline{E}_l^{(n)} < \bigwedge_{k=1,k\neq l}^{2n} \overline{E}_k^{(n)},\, n+1 \leq l \leq 2n\right\}}.$$

So L_{n+1} is the index of the minimum of $\{\overline{E}_k^{(n)}, 1 \leq k \leq 2n\}$ indicating the winner of the competition. Also, for $n \geq 1$, define the Bernoulli random variable

$$B_{n+1} := \mathbf{1}_{\left\{\bigwedge_{k=1}^{n} \overline{E}_k^{(n)} > \bigwedge_{k=n+1}^{2n} \overline{E}_k^{(n)}\right\}} = \mathbf{1}_{\{L_{n+1} > n\}},$$

and given $\mathcal{D}(n)$, we have

$$\mathcal{D}(n+1) = \mathcal{D}(n) + (1 - B_{n+1})\boldsymbol{e}^{\text{in}}_{L_{n+1}} + B_{n+1}\boldsymbol{e}^{\text{out}}_{L_{n+1}-n} + B_{n+1}\boldsymbol{e}^{\text{in}}_{n+1} + (1 - B_{n+1})\boldsymbol{e}^{\text{out}}_{n+1}. \quad (3.2.6)$$

This increments the L_{n+1}-st pair by $(1, 0)$ if $B_{n+1} = 0$ and the $(L_{n+1} - n)$-th pair by (0,1) if $B_{n+1} = 1$; the first case corresponds to an increase of in-degree and the second case to an increase of out-degree. The recursion also assigns to pair $n + 1$ either $(1, 0)$ or $(0, 1)$ depending on the case. This construction expresses $\mathcal{D}(n + 1)$ as a function of $\mathcal{D}(n)$ and something independent, namely $\{E^{(n)}_j, j = 1, \ldots, 2n\}$ and therefore the process $\{\mathcal{D}(n), n \geq 1\}$ is an $(\mathbb{N}^2_+)^\infty$-valued Markov chain. Also, because of the initialization (3.2.3), a simple induction argument applied to (3.2.6) gives the sum of the components satisfies

$$\sum_j D^{\text{in}}_j(n) = \sum_j D^{\text{out}}_j(n) = n, \quad n \geq 1. \quad (3.2.7)$$

Then using (3.2.1), (3.2.7) and standard calculations with exponential rv's, we have for $v \in [n]$,

$$\begin{aligned}
&\mathbf{P}\Big(\mathcal{D}(n+1) = \mathcal{D}(n) + \boldsymbol{e}^{\text{in}}_v + \boldsymbol{e}^{\text{out}}_{n+1}|\mathcal{D}(n)) = \mathbf{P}(L_{n+1} = v\Big|\mathcal{D}(n)\Big)\\
=&\mathbf{P}\left(\overline{E}^{(n)}_v < \bigwedge_{k=1,k\neq v}^{2n} \overline{E}^{(n)}_k \middle| \mathcal{D}(n)\right) = \frac{\alpha(D^{\text{in}}_v(n) + \delta_{\text{in}})}{(1 + \delta_{\text{in}})n}, \quad (3.2.8)
\end{aligned}$$

and likewise

$$\begin{aligned}
&\mathbf{P}\Big(\mathcal{D}(n+1) = \mathcal{D}(n) + \boldsymbol{e}^{\text{out}}_v + \boldsymbol{e}^{\text{in}}_{n+1}\Big|\mathcal{D}(n)\Big) = \mathbf{P}(L_{n+1} = n + v|\mathcal{D}(n))\\
=&\mathbf{P}\left(\overline{E}^{(n)}_{n+v} < \bigwedge_{k=1,k\neq n+v}^{2n} \overline{E}^{(n)}_k \middle| \mathcal{D}(n)\right) = \frac{\gamma(D^{\text{out}}_v(n) + \delta_{\text{out}})}{(1 + \delta_{\text{out}})n}. \quad (3.2.9)
\end{aligned}$$

These probabilities agree with the attachment probabilities in α- and γ-schemes, respectively.

3.2.2 Switched Birth Immigration Processes

In this section, we introduce a pair of switched birth immigration processes (SBI processes). This lays the foundation for Chapter 3.3, where we embed the in- and out-degree sequences of a fixed network node into a pair of SBI processes and derive the asymptotic limit of the degree growth.

Birth immigration processes.

We start with a brief review of the birth immigration process. A linear birth process with immigration (BI process), $\{Z(t) : t \geq 0\}$, having lifetime parameter $\lambda > 0$ and immigration parameter $\theta \geq 0$ is a continuous time Markov process with state space $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ and transition rate

$$q^Z_{k,k+1} = \lambda k + \theta, \qquad k \geq 0.$$

When $\theta = 0$ there is no immigration and the BI process becomes a pure birth process and in such cases, the process is usually taken to start at 1.

For $\theta > 0$, the BI process starting from 0 can be constructed from a Poisson process and an independent family of iid linear birth processes [57]. Suppose that $N_\theta(t)$ is the counting function of homogeneous Poisson points $0 < \tau_1 < \tau_2 < \ldots$ with rate θ and independent of this Poisson process we have independent copies of a linear birth process $\{\zeta_i(t) : t \geq 0\}_{i\geq 1}$ with parameter $\lambda > 0$ and $\zeta_i(0) = 1$ for $i \geq 1$. The BI process $Z(t), t \geq 0$ is a shot noise process with $Z(0) = 0$ and for $t \geq 0$,

$$Z(t) := \sum_{i=1}^{\infty} \zeta_i(t - \tau_i)\mathbf{1}_{\{t \geq \tau_i\}} = \sum_{i=1}^{N_\theta(t)} \zeta_i(t - \tau_i). \tag{3.2.10}$$

Theorem 3.2.1 modifies slightly the statement of [57, Theorem 5] summarizing the asymptotic behavior of the BI process. This is also reviewed in [68].

Theorem 3.2.1. *For* $\{Z(t) : t \geq 0\}$ *as in* (3.2.10), *we have as* $t \to \infty$,

$$e^{-\lambda t} Z(t) \xrightarrow{a.s.} \sum_{i=1}^{\infty} W_i e^{-\lambda \tau_i} =: \sigma \tag{3.2.11}$$

where $\{W_i : i \geq 1\}$ *are independent unit exponential random variables satisfying a.s. for each* $i \geq 1$,

$$W_i = \lim_{t \to \infty} e^{-t} \zeta_i(t).$$

The random variable σ *in* (3.2.11) *is a.s. finite and has a Gamma density given by*

$$f(x) = \frac{1}{\Gamma(\theta/\lambda)} x^{\theta/\lambda - 1} e^{-x}, \qquad x > 0.$$

Remark 3.2.2. For a BI process $\{Z'(t)\}_{t \geq 0}$ with $Z'(0) = j \geq 1$, modifying the representation in (3.2.10) gives

$$Z'(t) = \sum_{i=1}^{j} \zeta_i(t) + \sum_{i=j+1}^{\infty} \zeta_i(t - \tau_i) \mathbf{1}_{\{t \geq \tau_i\}}.$$

Therefore, $e^{-\lambda t} Z'(t) \xrightarrow{\text{a.s.}} \sigma'$ where σ' has a Gamma density given by $g(x) = x^{j + \theta/\lambda - 1} e^{-x} / \Gamma(j + \theta/\lambda)$, $x > 0$.

Switched birth immigration processes.

A switched birth immigration (SBI) process uses a Bernoulli choice variable to choose randomly from two independent BI processes with the same linear transition rates with one starting from 1 at $t = 0$ and the other starting from 0. A *pair* of SBI processes takes two SBI processes which are linked through the same Bernoulli choice variable.

Process	$I^{(0)}(t)$	$I^{(1)}(t)$	$O^{(0)}(t)$	$O^{(1)}(t)$
$t = 0$	0	1	1	0
Rate	$(1-p)(k+\delta_1)$		$p(k+\delta_2)$	

Table 3.2.1: Ingredients for a pair of switched BI processes.

Suppose that J is a Bernoulli switching random variable with

$$\mathbf{P}(J=1) = p = 1 - \mathbf{P}(J=0),$$

and $\{I^{(0)}(t) : t \geq 0\}$, $\{I^{(1)}(t) : t \geq 0\}$, $\{O^{(0)}(t) : t \geq 0\}$, $\{O^{(1)}(t) : t \geq 0\}$ are four independent BI processes (also independent of J) with $I^{(0)}(0) = O^{(1)}(0) = 0$, $I^{(1)}(0) = O^{(0)}(0) = 1$ and transition rates

$$q^{I^{(0)}}_{k,k+1} = (1-p)(k+\delta_1), \qquad q^{O^{(1)}}_{k,k+1} = p(k+\delta_2), \quad \text{for } k \geq 0,$$
$$q^{I^{(1)}}_{k,k+1} = (1-p)(k+\delta_1), \qquad q^{O^{(0)}}_{k,k+1} = p(k+\delta_2), \quad \text{for } k \geq 1,\ \delta_1, \delta_2 > 0.$$

See Table 3.2.1 for quick reminders. Then we construct a pair of SBI processes $\{(I^{(J)}(t), O^{(J)}(t)) : t \geq 0\}$ using five independent ingredients:

$$(I^{(J)}(t), O^{(J)}(t)) := (1-J)(I^{(0)}(t), O^{(0)}(t)) + J(I^{(1)}(t), O^{(1)}(t)), \qquad t \geq 0. \tag{3.2.12}$$

We then consider the convergence of the pair of SBI processes, $(e^{-(1-p)t} I^{(J)}(t), e^{-pt} O^{(J)}(t))$, as $t \to \infty$. Write a Gamma random variable X with density $f_X(x) = b^a x^{a-1} e^{-bx}/\Gamma(a)$, $x > 0$ and $a, b > 0$, as $X \sim \Gamma(a, b)$. Then from Theorem 3.2.1, Remark 3.2.2 and (3.2.12), we have with $X^{(0)}$, $Y^{(0)}$, $X^{(1)}$, $Y^{(1)}$ being four independent Gamma random variables and $X^{(0)} \sim \Gamma(\delta_0, 1)$, $Y^{(0)} \sim \Gamma(1+\delta_1, 1)$, $X^{(1)} \sim \Gamma(1+\delta_0, 1)$, $Y^{(1)} \sim \Gamma(\delta_1, 1)$, as $t \to \infty$,

$$(e^{-(1-p)t} I^{(J)}(t), e^{-pt} O^{(J)}(t)) \xrightarrow{\text{a.s.}} (1-J)(X^{(0)}, Y^{(0)}) + J(X^{(1)}, Y^{(1)}) =: (X^{(J)}, Y^{(J)}). \tag{3.2.13}$$

Also, $(X^{(J)}, Y^{(J)})$ has joint density

$$f_{X^{(J)},Y^{(J)}}(x,y) = (1-p)\,\frac{x^{\delta_0-1}e^{-x}}{\Gamma(\delta_0)}\frac{y^{\delta_1}e^{-y}}{\Gamma(1+\delta_1)} + p\,\frac{x^{\delta_0}e^{-x}}{\Gamma(1+\delta_0)}\frac{y^{\delta_1-1}e^{-y}}{\Gamma(\delta_1)}, \quad x, y > 0. \tag{3.2.14}$$

3.3 Embedding Process

In order to prove the weak convergence of the sequence of empirical measures in (3.1.6), we need to embed the in- and out-degree sequences $\{(D_v^{\text{in}}(n), D_v^{\text{out}}(n)), v \in [n], n \geq 1\}$ into a process constructed from pairs of SBI processes, as specified in Chapter 3.2.2. The embedding idea is proposed in [2] and has been used in [68] to model two different undirected linear preferential attachment models.

3.3.1 Embedding

Here we discuss how to embed the directed network growth model into a process constructed from an infinite sequence of SBI pairs.

Directed network model and SBI processes

The building blocks of the embedding procedure is an infinite family of independent BI processes

$$\left\{I_1(t), O_1(t), I_v^{(0)}(t), I_v^{(1)}(t), O_v^{(0)}(t), O_v^{(1)}(t) : v \geq 2, t \geq 0\right\},$$

defined on the same probability space and satisfying:

(i) $(I_1(0), O_1(0)) = 1$, $(I_v^{(0)}(0), O_v^{(0)}) = (0, 1)$ and $(I_v^{(1)}(0), O_v^{(1)}(0)) = (1, 0)$, for each $v \geq 2$.

(ii) Any process labeled with an I is a BI process with transition rates

$$q^I_{k,k+1} = \frac{c_{\text{in}}}{c_{\text{in}} + c_{\text{out}}}(k + \delta_{\text{in}}), \qquad \delta_{\text{in}} > 0,$$

and any process labeled with an O is a BI process with transition rates

$$q^O_{k,k+1} = \frac{c_{\text{out}}}{c_{\text{in}} + c_{\text{out}}}(k + \delta_{\text{out}}) \qquad \delta_{\text{out}} > 0.$$

These hold for $k \geq 0$ when $v \geq 2$ and $k \geq 1$ for I_1, O_1.

On $(\mathbb{N}^2)^\infty$, define

$$\mathbf{Z}^{(1)} = \{\mathbf{Z}^{(1)}_t : t \geq 0\} := \left\{\Big((I_1(t), O_1(t)), (0,0), \ldots\Big) : t \geq 0\right\}$$

and the σ-algebra $\mathcal{F}^{(1)}_t := \sigma\left\{\mathbf{Z}^{(1)}_t : 0 \leq s \leq t\right\}$ so that $\mathbf{Z}^{(1)}$ is strong Markov with respect to $\{\mathcal{F}^{(1)}_t\}$. Set $T_1 = 0$ and define the stopping time T_2 with respect to $\{\mathcal{F}^{(1)}_t, t \geq 0\}$ as

$$T_2 := \inf\left\{t \geq 0 : \mathbf{Z}^{(1)}_t \text{ jumps}\right\}. \tag{3.3.1}$$

Then T_2 is the minimum of two independent exponential r.v.'s with means

$$\left(\frac{c_{\text{in}}}{c_{\text{in}} + c_{\text{out}}}(1 + \delta_{\text{in}})\right)^{-1} \qquad \text{and} \qquad \left(\frac{c_{\text{out}}}{c_{\text{in}} + c_{\text{out}}}(1 + \delta_{\text{out}})\right)^{-1}.$$

From (3.2.1), we have

$$\mathbf{P}[T_2 > t] = e^{-(c_{\text{in}}+c_{\text{out}})^{-1}t}, \qquad t > 0.$$

Let $J_2 := \mathbf{1}_{\{O_1 \text{ jumps first}\}}$ so that $\mathbf{P}[J_2 = 1] = \gamma$. Also, let $\widetilde{L}_2$ be index of the (I, O)-pair that jumps first at T_2 which in this case is 1. However, note that $\left(\widetilde{L}_2, J_2\right)$ determines which one of I_1 and O_1 will jump at T_2, and T_2 is independent of $\left(\widetilde{L}_2, J_2\right)$ by the property of independent exponential r.v.'s (cf. [49, Exercise 4.45(a)]). In addition, we also have $T_2, \widetilde{L}_2, J_2 \in \mathcal{F}^{(1)}_{T_2}$, that is, measurable with respect to $\mathcal{F}^{(1)}_{T_2}$.

Now use the independent quantities $J_2, (I^{(0)}_2, O^{(0)}_2), (I^{(1)}_2, O^{(1)}_2)$ to define a pair of SBI processes $(I_2, O_2) = ((I^{(J_2)}_2, O^{(J_2)}_2)$ as in (3.2.12). Let $z_2(t) := ((0,0), (I^{(J_2)}_2(t), O^{(J_2)}_2(t)), (0,0), \ldots)$ and

$$\mathbf{Z}^{(2)} = \{\mathbf{Z}^{(2)}_t : t \geq 0\} := \left\{\mathbf{Z}^{(1)}_{t+T_2} + z_2(t) : t \geq 0\right\}.$$

Define the σ-algebra

$$\mathcal{F}^{(2)}_{t+T_2} := \sigma\left\{\mathbf{Z}^{(2)}_s : 0 \le s \le t\right\} \bigvee \mathcal{F}^{(1)}_{T_2},$$

so that $\mathbf{Z}^{(2)}$ is strong Markov with respect to $\{\mathcal{F}^{(2)}_{t+T_2}, t \ge 0\}$. Also, let

$$\tau_3 := \inf\left\{t \ge 0 : \mathbf{Z}^{(2)}_t \text{ jumps}\right\}, \qquad T_3 := T_2 + \tau_3,$$

and $J_3 := \mathbf{1}_{\left\{\text{One of } O_1(T_2+\cdot),\, O_2^{(J_2)}(\cdot) \text{ jumps first}\right\}}$. Denote the index of the (I,O)-pair that jumps at T_3 by $\widetilde{L}_3$ and write $\mathbf{P}^{\mathcal{F}^{(1)}_{T_2}}(\cdot) := \mathbf{P}(\cdot|\mathcal{F}^{(1)}_{T_2})$, $\mathbf{P}_z(\mathbf{Z}_t \in \cdot) := \mathbf{P}(\mathbf{Z}_t \in \cdot|\mathbf{Z}_0 = z)$. Then by the strong Markov property, we have

$$\mathbf{P}^{\mathcal{F}^{(1)}_{T_2}}\left(\mathbf{Z}^{(2)}_t \in \cdot\right) = \mathbf{P}_{\mathbf{Z}^{(1)}_{T_2}+z_2(0)}\left(\mathbf{Z}^{(1)}_t + z_2(t) \in \cdot\right).$$

Therefore, with respect to $\mathbf{P}^{\mathcal{F}^{(1)}_{T_2}}$, τ_3 is the minimum of 4 independent exponential r.v.'s with means $\left(\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(I_1(T_2)+\delta_{\text{in}})\right)^{-1}$, $\left(\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(O_1(T_2)+\delta_{\text{out}})\right)^{-1}$, $\left(\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(J_2+\delta_{\text{in}})\right)^{-1}$ and $\left(\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(1-J_2+\delta_{\text{out}})\right)^{-1}$. Note that $(I_1(T_2), O_1(T_2)) = (2 - J_2, 1 + J_2)$. We then have the following:

1. $\mathbf{P}^{\mathcal{F}^{(1)}_{T_2}}(\tau_3 > t) = e^{-2(c_{\text{in}}+c_{\text{out}})^{-1}t}$, $t > 0$.
2. $\mathbf{P}^{\mathcal{F}^{(1)}_{T_2}}(J_3 = 1) = \gamma$ and τ_3 is independent of $(\widetilde{L}_3, J_3)$ with respect to $\mathbf{P}^{\mathcal{F}^{(1)}_{T_2}}$.
3. The random variables $T_3, \widetilde{L}_3, J_3 \in \mathcal{F}^{(2)}_{T_3} = \mathcal{F}^{(2)}_{\tau_3+T_2}$.

Continue in this way to use the conditionally independent quantities J_3, $(I^{(0)}_3, O^{(0)}_3)$ and $(I^{(1)}_3, O^{(1)}_3)$ to define a pair of SBI processes $(I_3, O_3) = (I^{(J_3)}_3, O^{(J_3)}_3)$ as in (3.2.12). In general, for $n \ge 3$, set

$$\begin{aligned}\mathbf{Z}^{(n)}_t := \Big(&(I_1(T_n+t), O_1(T_n+t)), (I^{(J_2)}_2(T_n - T_2 + t), O^{(J_2)}_2(T_n - T_2 + t)),\\ &\ldots, (I^{(J_n)}_n(t), O^{(J_n)}_n(t)), (0,0), \ldots\Big), \qquad t \ge 0,\end{aligned}$$

$\mathcal{F}^{(n)}_{t+T_n} := \sigma\left\{\mathbf{Z}^{(n)}_s : 0 \le s \le t\right\} \bigvee \mathcal{F}^{(n-1)}_{T_n}$, $\tau_{n+1} := \inf\{t \ge 0 : \mathbf{Z}^{(n)}_t \text{ jumps}\}$ and $T_{n+1} := T_n + \tau_{n+1}$. Also, define

- $J_{n+1} := \mathbf{1}_{\left\{\text{One of } O_1(T_n+\cdot),\, O_k^{(J_k)}(T_n - T_k + \cdot),\, k=2,\ldots,n \text{ jumps first}\right\}}$, and
- $\widetilde{L}_{n+1}$ is the index of the (I, O)-pair that jumps first among $(I_1(T_n + t), O_1(T_n + t)), (I_k(T_n - T_k + t, O_k(T_n - T_k + t),\ k = 2, \ldots, n.$

Note that with

$$z_n(t) := \left((0,0), \ldots, \underbrace{\left(I_n^{(J_n)}(t), O_n^{(J_n)}(t)\right)}_{n\text{-th pair}}, (0,0), \ldots \right),$$

we have $\mathbf{Z}_t^{(n)} = \mathbf{Z}_{\tau_n+t}^{(n-1)} + z_n(t)$. Using the strong Markov property gives

$$\mathbf{P}^{\mathcal{F}_{T_n}^{(n-1)}}\left(\mathbf{Z}_t^{(n)} \in \cdot\right) = \mathbf{P}_{\mathbf{Z}_{\tau_n}^{(n-1)}+z_n(0)}\left(\mathbf{Z}_t^{(1)} + \sum_{k=2}^{n} z_k(t) \in \cdot\right).$$

Then with respect to $\mathcal{F}_{T_n}^{(n-1)}$, τ_{n+1} is the minimum of $2n$ independent exponential r.v.'s with means

$$\left(\frac{c_{\text{in}}}{c_{\text{in}} + c_{\text{out}}}(I_1(T_n) + \delta_{\text{in}})\right)^{-1}, \left(\frac{c_{\text{out}}}{c_{\text{in}} + c_{\text{out}}}(O_1(T_n) + \delta_{\text{out}})\right)^{-1},$$
$$\left(\frac{c_{\text{in}}}{c_{\text{in}} + c_{\text{out}}}(I_k^{(J_k)}(T_n - T_k) + \delta_{\text{in}})\right)^{-1}, \left(\frac{c_{\text{out}}}{c_{\text{in}} + c_{\text{out}}}(O_k^{(J_k)}(T_n - T_k) + \delta_{\text{out}})\right)^{-1},\ k = 2, \ldots, n.$$

This implies:

1. The random variable τ_{n+1} is independent of $(\widetilde{L}_{n+1}, J_{n+1})$ with respect to $\mathbf{P}^{\mathcal{F}_{T_n}^{(n-1)}}$.
2. The random variables $T_{n+1}, \widetilde{L}_{n+1}, J_{n+1} \in \mathcal{F}_{T_{n+1}}^{(n)}$.

Set $\tau_2 := T_2$. Then from this construction follow properties of the distribution of $\{\tau_n\}_{n\geq 2}$ and $\{J_n\}_{n\geq 2}$.

Lemma 3.3.1. *Suppose $\{T_n\}_{n\geq 1}$, $\{\tau_n\}_{n\geq 2}$ and $\{J_n\}_{n\geq 2}$ are defined as above. Then:*

(i) The sequence $\{J_n\}$ is independent of $\{\tau_n\}$.

(ii) The sequence $\{J_n\}$ is a sequence of iid Bernoulli random variables with

$$\boldsymbol{P}(J_n = 1) = \gamma = 1 - \boldsymbol{P}(J_n = 0), \qquad n \geq 2. \tag{3.3.2}$$

(iii) The sequence $\{\tau_n\}_{n\geq 2}$ satisfies

$$\{\tau_{n+1} : n \geq 1\} \stackrel{d}{=} \left\{ \frac{E_n}{(c_{in} + c_{out})^{-1} n}, n \geq 1 \right\}, \tag{3.3.3}$$

where $\{E_n : n \geq 1\}$ is a sequence of iid unit exponential random variables. So $\{T_n\}$ are the birth times of a linear birth process with birth rate $(c_{in} + c_{out})^{-1}$.

Proof. For brevity of notation, write $\lambda_n^{I_1} = \frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(I_1(T_n)+\delta_{\text{in}})$, $\lambda_n^{O_1} = \frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(O_1(T_n)+\delta_{\text{out}})$ and for $2 \leq k \leq n, n \geq 2$,

$$\begin{aligned}
\lambda_n^{I_k} &= \frac{c_{\text{in}}}{c_{\text{in}} + c_{\text{out}}}(I_k^{(J_k)}(T_n - T_k) + \delta_{\text{in}}), \\
\lambda_n^{O_k} &= \frac{c_{\text{out}}}{c_{\text{in}} + c_{\text{out}}}(O_k^{(J_k)}(T_n - T_k) + \delta_{\text{out}}).
\end{aligned}$$

At each T_n, $n \geq 2$, we start a new pair of SBI processes $(I_n(\cdot), O_n(\cdot))$ with initial value $(J_n, 1 - J_n)$ and one of $(I_k(\cdot), O_k(\cdot))$, $1 \leq k \leq n-1$ increases by $(1 - J_n, J_n)$. This corresponds in the network, for instance if $J_n = 1$, to one of the existing $n - 1$ nodes having an out-degree increase by 1 and a new node n with in-degree 1 and out-degree 0. Therefore (cf. (3.2.7)),

$$I_1(T_n) + \sum_{k=2}^{n} I_k^{(J_k)}(T_n - T_k) = O_1(T_n) + \sum_{k=2}^{n} O_k^{(J_k)}(T_n - T_k) = n. \tag{3.3.4}$$

Hence, for $n \geq 2$, $t_l > 0$ and $j_l \in \{0, 1\}$ for $l = 2, \ldots, n + 1$,

$$\begin{aligned}
\mathbf{P}\left(\bigcap_{l=2}^{n+1}[\tau_l > t_l, J_l = j_l]\right) &= \mathbf{E}\left[\mathbf{P}^{\mathcal{F}_{T_n}^{(n-1)}}\left(\tau_{n+1} > t_{n+1}, J_{n+1} = j_{n+1}, \bigcap_{l=2}^{n}\{\tau_l > t_l, J_l = j_l\right)\right] \\
&= \mathbf{E}\left[\mathbf{1}_{\bigcap_{l=2}^{n}\{\tau_l > t_l, J_l = j_l\}}\mathbf{P}^{\mathcal{F}_{T_n}^{(n-1)}}(\tau_{n+1} > t_{n+1}, J_{n+1} = j_{n+1})\right],
\end{aligned} \tag{3.3.5}$$

since $(\tau_l, J_l, l = 2, \ldots, n) \in \mathcal{F}_{T_n}^{(n-1)}$. Also, we know that with respect to $\mathbf{P}^{\mathcal{F}_{T_n}^{(n-1)}}$, τ_{n+1} is the minimum of $2n$ independent exponential r.v.'s and J_{n+1} is independent of τ_{n+1}. Therefore,

$$\mathbf{P}^{\mathcal{F}_{T_n}^{(n-1)}}(\tau_{n+1} > t_{n+1}, J_{n+1} = j_{n+1}) = \mathbf{P}^{\mathcal{F}_{T_n}^{(n-1)}}(\tau_{n+1} > t_{n+1})\, \mathbf{P}^{\mathcal{F}_{T_n}^{(n-1)}}(J_{n+1} = j_{n+1}). \tag{3.3.6}$$

Note that

$$\begin{aligned}\mathbf{P}^{\mathcal{F}_{T_n}^{(n-1)}}(\tau_{n+1} > t_{n+1}) &= \exp\left\{-t_{n+1}\sum_{k=1}^{n}\left(\lambda_n^{I_k} + \lambda_n^{O_k}\right)\right\} \\ &= \exp\left\{-t_{n+1}(c_{\text{in}} + c_{\text{out}})^{-1}n\right\},\end{aligned} \tag{3.3.7}$$

and assuming $j_{n+1} = 1$, we have

$$\mathbf{P}^{\mathcal{F}_{T_n}^{(n-1)}}(J_{n+1} = 1) = \frac{\sum_{k=1}^{n}\lambda_n^{O_k}}{\sum_{k=1}^{n}(\lambda_n^{I_k} + \lambda_n^{O_k})} = \gamma. \tag{3.3.8}$$

So (3.3.5) becomes (continuing to suppose $j_{n+1} = 1$),

$$\mathbf{P}\left(\bigcap_{l=2}^{n+1}[\tau_l > t_l, J_l = j_l]\right) = \gamma \exp\left\{-t_{n+1}(c_{\text{in}} + c_{\text{out}})^{-1}n\right\}\mathbf{P}\left(\bigcap_{l=2}^{n}[\tau_l > t_l, J_l = j_l]\right).$$

If $j_{n+1} = 0$, γ is replaced by α on the right side. This is sufficient for the proof of the Lemma. □

Embedding

The following embedding theorem is similar to those proved in [2, 68] and summarizes how to embed in the paired SBI process constructions.

Theorem 3.3.2. *Suppose that $\{T_n\}_{n\geq 1}$ and $\{\mathbf{Z}_t^{(n)} : t \geq 0\}$ are as defined in Chapter 3.3.1. Then in $((\mathbb{N}^2)^\infty)^\infty$,*

$$\{\mathcal{D}(n), n \geq 1\} \stackrel{d}{=} \left\{\mathbf{Z}_0^{(n)}, n \geq 1\right\}.$$

Proof. The proof relies on both $\{\mathcal{D}(n), n \geq 1\}$ and $\{\boldsymbol{Z}_0^{(n)}, n \geq 1\}$ being Markov chains with the same transition probabilities. It is similar to that of [2, Theorem 2.1] and [68, Theorem 2] which we now outline.

Define

$$\widetilde{\boldsymbol{d}}_j^{(J_n)} := \left((0,0), \ldots, \underbrace{(1-J_n, J_n)}_{j\text{-th pair}}, (0,0), \ldots, (0,0), \underbrace{(J_n, 1-J_n)}_{n\text{-th pair}}, (0,0), \ldots \right)$$

Recall that $\widetilde{L}_{n+1}$ is the index of the (I, O)-pair that jumps at T_{n+1}. Then we have

$$\boldsymbol{Z}_0^{(n+1)} = \boldsymbol{Z}_0^{(n)} + \widetilde{\boldsymbol{d}}_{\widetilde{L}_{n+1}}^{(J_{n+1})}. \tag{3.3.9}$$

This expresses $\boldsymbol{Z}_0^{(n+1)}$ as a function of $\mathcal{F}_{T_n}^{(n-1)}$-measurable random elements and random elements independent of $\mathcal{F}_{T_n}^{(n-1)}$, namely:

1. $\boldsymbol{Z}_0^{(n)} \in \mathcal{F}_{T_n}^{(n-1)}$;
2. J_{n+1} which is independent of $\mathcal{F}_{T_n}^{(n-1)}$ (by Lemma 3.3.1; see (3.3.8));
3. $\widetilde{L}_{n+1}$ which is a function of $(\lambda_n^{I_k} + \lambda_n^{O_k}, k = 2, \ldots, n) \in \mathcal{F}_{T_n}^{(n-1)}$ and conditionally on $\mathcal{F}_{T_n}^{(n-1)}$, $2n$ i.i.d exponential r.v.s which are independent of $\mathcal{F}_{T_n}^{(n-1)}$.

Hence, both $\{\mathcal{D}(n), n \geq 1\}$ and $\{\boldsymbol{Z}_0^{(n)}, n \geq 1\}$ are Markov on the state space $(\mathbb{N}^2)^\infty$.

When $n = 1$,

$$\begin{aligned}\boldsymbol{Z}_0^{(1)} &= \Big((I_1(0), O_1(0)), (0,0), \ldots\Big) = ((1,1), (0,0), \ldots) \\ &= \Big((D_1^{\text{in}}(1), D_1^{\text{out}}(1)), (0,0), \ldots\Big) = \mathcal{D}(1),\end{aligned}$$

so to prove equality in distribution for any n, it suffices to verify that the transition probability from $\boldsymbol{Z}_0^{(n)}$ to $\boldsymbol{Z}_0^{(n+1)}$ is the same as that from $\mathcal{D}(n)$ to $\mathcal{D}(n+1)$ which is given in (3.2.8) and (3.2.9). In the SBI setup, applying Lemma 3.3.1 gives for any $2 \le v \le n$,

$$\begin{aligned}
\mathbf{P}^{\mathcal{F}^{(n-1)}_{T_n}}\Big(\boldsymbol{Z}_0^{(n+1)} =& \boldsymbol{Z}_0^{(n)} + \boldsymbol{e}_v^{\text{in}} + \boldsymbol{e}_{n+1}^{\text{out}})\Big) = \mathbf{P}^{\mathcal{F}^{(n-1)}_{T_n}}\Big(J_{n+1} = 0, \widetilde{L}_{n+1} = v\Big) \\
&= \frac{\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(I_v^{(J_v)}(T_n - T_v) + \delta_{\text{in}})}{(c_{\text{in}} + c_{\text{out}})^{-1} n} = \alpha \frac{I_v^{(J_v)}(T_n - T_v) + \delta_{\text{in}}}{(1+\delta_{\text{in}})n}, \\
\mathbf{P}^{\mathcal{F}^{(n-1)}_{T_n}}\Big(\boldsymbol{Z}_0^{(n+1)} =& \boldsymbol{Z}_0^{(n)} + \boldsymbol{e}_{n+1}^{\text{in}} + \boldsymbol{e}_v^{\text{out}})\Big) = \mathbf{P}^{\mathcal{F}^{(n-1)}_{T_n}}\Big(J_{n+1} = 1, \widetilde{L}_{n+1} = v\Big) \\
&= \frac{\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(O_v^{(J_v)}(T_n - T_v) + \delta_{\text{out}})}{(c_{\text{in}} + c_{\text{out}})^{-1} n} = \gamma \frac{O_v^{(J_v)}(T_n - T_v) + \delta_{\text{out}}}{(1+\delta_{\text{out}})n}.
\end{aligned}$$

For $2 \le v \le n$, this agrees with the transition probabilities in (3.2.8) and (3.2.9) respectively; the case for $v = 1$ is similar. □

3.3.2 Asymptotic properties

With the embedding technique specified in Chapter 3.3.1, the asymptotic behavior of the in- and out-degree growth in a preferential attachment model can be characterized explicitly. These asymptotic properties then help us derive weak convergence of the empirical measure. For brevity of notation, we will write $I_v^{(J_v)}$, $O_v^{(J_v)}$ as I_v, O_v, $v \ge 2$, in the rest of this chapter.

Convergence of the in- and out-degrees for a fixed node.

We first consider the asymptotic behavior of the in- and out-degrees for a fixed node, i.e. $(D_v^{\text{in}}(n), D_v^{\text{out}}(n))$ for a fixed v. To do this, we make use of the embedding results in Theorem 3.3.2, which translates the convergence of the degrees to the setting of $\{(I_v(t - T_v), O_v(t - T_v)) : t \ge T_v\}_{1 \le v \le n}$. Results are summarized in Theorem 3.3.3.

Theorem 3.3.3. *Suppose that* $\{T_n : n \geq 1\}$ *and* $\{J_n : n \geq 2\}$ *are as defined in Chapter 3.3.1. Then:*

(i) The birth times $\{T_n\}_{n\geq 1}$ *satisfy that as* $n \to \infty$,

$$n \cdot e^{-(c_{in}+c_{out})^{-1}T_n} \xrightarrow{a.s.} W \qquad \text{and} \qquad W \sim Exp(1). \tag{3.3.10}$$

(ii) Let $(\sigma_1^{in}, \sigma_1^{out})$ *be a pair of independent Gamma random variables with densities*

$$f_{\sigma_1^{in}}(x) = \frac{x^{\delta_{in}}e^{-x}}{\Gamma(1+\delta_{in})} \quad \text{and} \quad f_{\sigma_1^{out}}(x) = \frac{x^{\delta_{out}}e^{-x}}{\Gamma(1+\delta_{out})},\ x > 0,\ \text{respectively,}$$

and for each $v \geq 2$, $(\sigma_v^{in}, \sigma_v^{out})$ *have joint density*

$$f_{(\sigma_v^{in},\sigma_v^{out})}(x,y) = \alpha\frac{x^{\delta_{in}-1}e^{-x}}{\Gamma(\delta_{in})}\frac{y^{\delta_{out}}e^{-y}}{\Gamma(1+\delta_{out})} + \gamma\frac{x^{\delta_{in}}e^{-x}}{\Gamma(1+\delta_{in})}\frac{y^{\delta_{out}-1}e^{-y}}{\Gamma(\delta_{out})}, \quad x,y > 0. \tag{3.3.11}$$

Then for a fixed $v \geq 1$, *we have, with* W *defined as in* (3.3.10),

$$\left(\frac{D_v^{in}(n)}{n^{c_{in}}}, \frac{D_v^{out}(n)}{n^{c_{out}}}\right) \Rightarrow \left(\frac{\sigma_v^{in}e^{-\frac{c_{in}}{c_{in}+c_{out}}T_v}}{W^{c_{in}}}, \frac{\sigma_v^{out}e^{-\frac{c_{out}}{c_{in}+c_{out}}T_v}}{W^{c_{out}}}\right) \quad n \to \infty. \tag{3.3.12}$$

Also, setting $D_v^{in}(n) = 0 = D_v^{out}(n)$ *for all* $v \geq n+1$, *we get as* $n \to \infty$,

$$\left(\max_{v\geq 1}\frac{D_v^{in}(n)}{n^{c_{in}}}, \max_{v\geq 1}\frac{D_v^{out}(n)}{n^{c_{out}}}\right) \Rightarrow \left(\max_{v\geq 1}\frac{\sigma_v^{in}e^{-\frac{c_{in}}{c_{in}+c_{out}}T_v}}{W^{c_{in}}}, \max_{v\geq 1}\frac{\sigma_v^{out}e^{-\frac{c_{out}}{c_{in}+c_{out}}T_v}}{W^{c_{out}}}\right). \tag{3.3.13}$$

Here T_v, $(\sigma_v^{in}, \sigma_v^{out})$ *and* W *are independent for all* $v \geq 2$.

Remark 3.3.4. According to the embedding results in Theorem 3.3.2, (3.3.12) also implies that there exists random variables $D_v^{(1)}$, $D_v^{(2)}$, $v \geq 1$, on the space of $(D_v^{\text{in}}(n), D_v^{\text{out}}(n))_{v\geq 1}$ satisfying $D_v^{(1)} \stackrel{d}{=} W^{-c_{\text{in}}}\sigma_v^{\text{in}}e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}T_v}$ and $D_v^{(2)} \stackrel{d}{=} W^{-c_{\text{out}}}\sigma_v^{\text{out}}e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}T_v}$, $v \geq 1$, such that as $n \to \infty$,

$$\left(\frac{D_v^{\text{in}}(n)}{n^{c_{\text{in}}}}, \frac{D_v^{\text{out}}(n)}{n^{c_{\text{out}}}}\right) \xrightarrow{\text{a.s.}} \left(D_v^{(1)}, D_v^{(2)}\right).$$

Proof. (i) From Lemma 3.3.1(i), $\{T_n : n \geq 1\}$ are jump times of a pure birth process starting from 1 and transition rate

$$q_{j,j+1} = (c_{\text{in}} + c_{\text{out}})^{-1} j, \qquad j \geq 1.$$

Therefore, (3.3.10) follows from applying the known convergence results of linear birth processes; see [49, Theorem 5.11.4] and [33, 69], among other sources.

(ii) By Theorem 3.3.2, to show (3.3.12), it suffices to show that as $n \to \infty$,

$$\left(\frac{I_v(T_n - T_v)}{n^{c_{\text{in}}}}, \frac{O_v(T_n - T_v)}{n^{c_{\text{out}}}} \right) \xrightarrow{\text{a.s.}} \left(\frac{\sigma_v^{\text{in}} e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}} T_v}}{W^{c_{\text{in}}}}, \frac{\sigma_v^{\text{out}} e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}} T_v}}{W^{c_{\text{out}}}} \right),. \tag{3.3.14}$$

With (3.3.10) available, we prove (3.3.14) by showing the convergence of

$$\left(e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(t-T_v)} I_v(t - T_v), e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(t-T_v)} O_v(t - T_v) \right),$$

as $t \to \infty$. According to the construction of the processes $\{(I_v(t - T_v), O_v(t - T_v) : t \geq T_v)\}_{v \geq 1}$, we know that $(I_1(0), O_1(0)) = (1, 1)$. Then applying the convergence result of a BI process in Remark 3.2.2, we have for independent $(\sigma_1^{\text{in}}, \sigma_1^{\text{out}}) \sim (\Gamma(1 + \delta_{\text{in}}, 1), \Gamma(1 + \delta_{\text{out}}, 1))$,

$$\left(e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}} t} I_1(t), e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}} t} O_v(t) \right) \xrightarrow{\text{a.s.}} (\sigma_1^{\text{in}}, \sigma_1^{\text{out}}), \qquad t \to \infty.$$

Moreover, it follows from (3.2.13) and (3.2.14) that

$$\left(e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(t-T_v)} I_v(t - T_v), e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(t-T_v)} O_v(t - T_v) \right) \xrightarrow{\text{a.s.}} (\sigma_v^{\text{in}}, \sigma_v^{\text{out}}), \qquad t \to \infty, \tag{3.3.15}$$

with σ_v^{in} and σ_v^{out} having the joint density as in (3.3.11).

Replacing t with T_n in (3.3.15) gives

$$\left(\frac{I_v(T_n - T_v)}{e^{\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}} T_n}}, \frac{O_v(T_n - T_v)}{e^{\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}} T_n}} \right) \xrightarrow{\text{a.s.}} \left(\sigma_v^{\text{in}} e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}} T_v}, \sigma_v^{\text{out}} e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}} T_v} \right), \quad \text{as } n \to \infty. \tag{3.3.16}$$

Therefore, combining (3.3.10) and (3.3.16) gives (3.3.12). For $v \geq 2$, the independence of $(\sigma_v^{\text{in}}, \sigma_v^{\text{out}})$ and T_v follows from the construction and the independence from W follows from [49, p. 443]; this completes the proof of (3.3.14).

(iii) We verify (3.3.13) by showing that as $n \to \infty$,

$$\left(\max_{v\geq 1} \frac{I_v(T_n - T_v)}{e^{\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}T_n}}, \max_{v\geq 1} \frac{O_v(T_n - T_v)}{e^{\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}T_n}}\right) \xrightarrow{\text{a.s.}} \left(\max_{v\geq 1} \sigma_v^{\text{in}} e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}T_v}, \max_{v\geq 1} \sigma_v^{\text{out}} e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}T_v}\right). \tag{3.3.17}$$

Then combining (3.3.17) with (3.3.10) gives the result. We use the proof machinery in [2, Proposition 3.1] to show (3.3.17), which is summarized in the following lemma.

Lemma 3.3.5. *Let* $a_{n,i} : 1 \leq i \leq n_{n\geq 1}$ *be a double array of non-negative numbers such that*

(1) For all $i \geq 1$, $\lim_{n\to\infty} a_{n,i} = a_i < \infty$,

(2) $\sup_{n\geq 1} a_{n,i} \leq b_i < \infty$ *and*

(3) $\lim_{i\to\infty} b_i = 0$.

Then $\max_{1\leq i\leq n} a_{n,i} \to \max_{i\geq 1} a_i$, *as* $n \to \infty$.

First note that for each $v \geq 1$,

$$I_v(T_n - T_v)e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(T_n - T_v)} \leq \sup_{t\geq 0} I_v(t)e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}t} =: \widetilde{I}_v,$$
$$O_v(T_n - T_v)e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(T_n - T_v)} \leq \sup_{t\geq 0} O_v(t)e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}t} =: \widetilde{O}_v.$$

Let $a^I_{n,v} := I_v(T_n - T_v)e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}T_n}$, $a^O_{n,v} := O_v(T_n - T_v)e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}T_n}$ for $1 \leq v \leq n$, and $b^I_v := \widetilde{I}_v e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}T_v}$, $b^O_v := \widetilde{O}_v e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}T_v}$ for $v \geq 1$. Then Lemma 3.3.5(1) is satisfied by

(3.3.16). Also, for each $v \geq 1$, $\sup_{n\geq 1} a^I_{n,v} \leq b^I_v$ and $\sup_{n\geq 1} a^O_{n,v} \leq b^O_v$, which satisfies the criterion in Lemma 3.3.5(2).

Following the proof of [2, Theorem 1.1], we check the condition in Lemma 3.3.5(3) by proving the claim that almost surely, for all $\epsilon > 0$,

$$\widetilde{I}_v \leq \epsilon v^{c_{\text{in}}}, \quad \text{and} \quad \widetilde{O}_v \leq \epsilon v^{c_{\text{out}}}, \quad \text{for all large } v. \tag{3.3.18}$$

Then as ϵ is arbitrary, it follows from (3.3.10) that $b^I_v \to 0$ and $b^O_v \to 0$ a.s. as $v \to \infty$. This completes checking the three criteria in Lemma 3.3.5 and therefore leads to (3.3.13).

To show (3.3.18), we use Markov's inequality: for any $r, r' > 0$ and $v \geq 2$,

$$\mathbf{P}(\widetilde{I}_v \geq \epsilon v^{c_{\text{in}}}) \leq \mathbf{E}(\widetilde{I}_2^r)/(\epsilon^r v^{r c_{\text{in}}}),$$
$$\mathbf{P}(\widetilde{O}_v \geq \epsilon v^{c_{\text{out}}}) \leq \mathbf{E}(\widetilde{O}_2^{r'})/(\epsilon^{r'} v^{r' c_{\text{out}}}),$$

since I_v, O_v, $v \geq 2$ are iid SBI processes. Hence, if we have

$$\mathbf{E}(\widetilde{I}_2^r) < \infty \ \text{ and } \ \mathbf{E}(\widetilde{O}_2^r) < \infty, \qquad \text{for } \ r > c_{\text{in}}^{-1}, r' > c_{\text{out}}^{-1}, \text{ respectively,} \tag{3.3.19}$$

then by Borel-Cantelli, the claim in (3.3.18) is justified. To prove (3.3.19), let

$$\widetilde{I^{(0)}}_2 := \sup_{t\geq 0} I_2^{(0)}(t) e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}} t}, \qquad \widetilde{I^{(1)}}_2 := \sup_{t\geq 0} I_2^{(1)}(t) e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}} t},$$
$$\widetilde{O^{(0)}}_2 := \sup_{t\geq 0} O_2^{(0)}(t) e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}} t}, \qquad \widetilde{O^{(1)}}_2 := \sup_{t\geq 0} O_2^{(1)}(t) e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}} t},$$

then by the construction of $(I_2(\cdot), O_2(\cdot))$, we have

$$\mathbf{E}(\widetilde{I}_2^r) = \alpha \mathbf{E}(\widetilde{I^{(0)}}_2^r) + \gamma \mathbf{E}(\widetilde{I^{(1)}}_2^r) < \infty,$$
$$\mathbf{E}(\widetilde{I}_2^{r'}) = \alpha \mathbf{E}(\widetilde{O^{(0)}}_2^{r'}) + \gamma \mathbf{E}(\widetilde{O^{(1)}}_2^{r'}) < \infty,$$

using the assumption that $I_2^{(0)}$, $I_2^{(1)}$, $O_2^{(0)}$ and $O_2^{(1)}$ are independent BI processes so that results in [2, Proposition 2.6] are still applicable here. This completes the proof of (3.3.17). $\square$

3.4 Convergence Results on Joint Degree Distributions

3.4.1 Convergence of the joint degree counts

Now we analyze the convergence of the joint empirical distribution of the in- and out-degrees $\{(D_v^{\text{in}}(n), D_v^{\text{out}}(n)) : v \in [n]\}$, using the SBI embedding technique. Let $B(a, p)$ be a negative binomial integer valued random variable with parameters $a > 0$ and $p \in (0, 1)$ (abbreviated as $NB(a, p)$), and the generating function of $B(a, p)$ is

$$\mathbf{E}\left(s^{B(a,p)}\right) = p^a(1 - (1 - p)s)^{-a}, \qquad 0 \leq s \leq 1.$$

We also use the notation $B(a, Z)$ to represent a r.v. having a mixture distribution such that the second parameter of the negative binomial r.v. is randomized by an independent r.v. Z.

Theorem 3.4.1. *Let $N_{i,j}(n)$ be the number of nodes with in-degree i and out-degree j in graph $G(n)$, then we have*

$$\frac{N_{i,j}(n)}{n} \xrightarrow{p} \boldsymbol{P}((\mathcal{I}, \mathcal{O}) = (i, j)), \qquad \text{as } n \to \infty. \tag{3.4.1}$$

The limit pair $(\mathcal{I}, \mathcal{O})$ can be represented in distribution as:

$$(\mathcal{I}, \mathcal{O}) \stackrel{d}{=} (1 - J)(X_1, 1 + Y_1) + J(1 + X_2, Y_2), \tag{3.4.2}$$

where

(i) J is a Bernoulli switching variable with $\boldsymbol{P}(J = 1) = 1 - \boldsymbol{P}(J = 0) = \gamma$.

(ii) Suppose $\{B^{(1)}(\delta_1, p) : p \in (0, 1)\}$, $\{B^{(2)}(\delta_1', p) : p \in (0, 1)\}$, $\{\widetilde{B}^{(1)}(\delta_2, p) : p \in (0, 1)\}$ and $\{\widetilde{B}^{(2)}(\delta_2', p) : p \in (0, 1)\}$, $\delta_1, \delta_1', \delta_2, \delta_2' > 0$, are four independent families of negative binomial variables, then

$$(X_1, Y_1) = \left(B^{(1)}\left(\delta_{in}, e^{-c_{in}T}\right), \widetilde{B}^{(1)}\left(1 + \delta_{out}, e^{-c_{out}T}\right)\right), \tag{3.4.3a}$$

$$(X_2, Y_2) = \left(B^{(2)}\left(1 + \delta_{in}, e^{-c_{in}T}\right), \widetilde{B}^{(2)}\left(\delta_{out}, e^{-c_{out}T}\right)\right), \tag{3.4.3b}$$

with T being an exponential random variable with unit mean, independent of J, $B^{(1)}$, $B^{(2)}$, $\widetilde{B}^{(1)}$ and $\widetilde{B}^{(2)}$.

Remark 3.4.2. Theorem 3.4.1 coincides with the known results proven in [52, 55], since $e^{c_{\text{in}}T}$ is a Pareto random variable on $[1, \infty)$ with index c_{in}^{-1}, denoted by Z, and $e^{c_{\text{out}}T} = Z^a$, with $a := c_{\text{out}}/c_{\text{in}}$.

Proof. The proof of [66, Lemma 3.1] verifies that

$$\left|\frac{N_{i,j}(n)}{n} - \frac{\mathbf{E}(N_{i,j}(n))}{n}\right| \xrightarrow{p} 0, \qquad \text{as } n \to \infty.$$

Hence, we are left to examine the difference $|\mathbf{E}(N_{i,j}(n))/n - \mathbf{P}((\mathcal{I}, \mathcal{O}) = (i, j))|$. By the embedding results in Theorem 3.3.2, we have

$$\begin{aligned}\frac{\mathbf{E}(N_{i,j}(n))}{n} &= \mathbf{E}\left\{\frac{1}{n}\sum_{v\in[n]} \mathbf{1}_{\left\{\left(D_v^{\text{in}}(n), D_v^{\text{out}}(n)\right)=(i,j)\right\}}\right\} = \frac{1}{n}\sum_{v\in[n]} \mathbf{P}\Big((D_v^{\text{in}}(n), D_v^{\text{out}}(n)) = (i, j)\Big) \\ &= \frac{1}{n}\sum_{v=1}^{n} \mathbf{P}\left[(I_v(T_n - T_v), O_v(T_n - T_v)) = (i, j)\right].\end{aligned} \tag{3.4.4}$$

Suppose that $\{B_v^{(1)}(\delta_{\text{in}}, p) : v \geq 1\}$, $\{B_v^{(2)}(1+\delta_{\text{in}}, p) : v \geq 1\}$, $\{\widetilde{B}_v^{(1)}(1+\delta_{\text{out}}, p) : v \geq 1\}$ and $\{\widetilde{B}_v^{(2)}(\delta_{\text{out}}, p) : v \geq 1\}$ are four independent sequences of negative binomial r.v.'s with given parameters. Then by the distribution of a BI process (cf. [57, Equation (2.2)] and [22, Theorem 3.11]), we have for any $v \geq 2$, $t \geq 0$ and $k \geq 0$,

$$\mathbf{P}(I_v^{(0)}(t) = k) = \mathbf{P}\left[B_v^{(1)}\left(\delta_{\text{in}}, e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}t}\right) = k\right], \tag{3.4.5a}$$

$$\mathbf{P}(I_v^{(1)}(t) = k) = \mathbf{P}\left[1 + B_v^{(2)}\left(1 + \delta_{\text{in}}, e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}t}\right) = k\right], \tag{3.4.5b}$$

$$\mathbf{P}(O_v^{(0)}(t) = k) = \mathbf{P}\left[1 + \widetilde{B}_v^{(1)}\left(1 + \delta_{\text{out}}, e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}t}\right) = k\right], \tag{3.4.5c}$$

$$\mathbf{P}(O_v^{(1)}(t) = k) = \mathbf{P}\left[\widetilde{B}_v^{(2)}\left(\delta_{\text{out}}, e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}t}\right) = k\right], \tag{3.4.5d}$$

and note the quantities on the right do not depend on v. Also, recall that $(I_v(t), O_v(t))_{v\geq 2}$, $t \geq 0$, are identically distributed such that,

$$I_v(t) = (1 - J_v)I_v^{(0)}(t) + J_v I_v^{(1)}(t), \quad O_v(t) = (1 - J_v)O_v^{(0)}(t) + J_v O_v^{(1)}(t).$$

Since for $v \geq 2$, the processes $I_v^{(0)}$, $I_v^{(1)}$, $O_v^{(0)}$ and $O_v^{(1)}$ are independent from each other, we then define for any $v \geq 2$,

$$\begin{aligned}\mathcal{B}_v^{(n)} := ((1 - J_v)B_v^{(1)}(\delta_{\text{in}}, e^{-(T_n - T_v)}) + J_v(1 + B_v^{(2)}(1 + \delta_{\text{in}}, e^{-(T_n - T_v)}),\\ (1 - J_v)(1 + \widetilde{B}_v^{(1)}(1 + \delta_{\text{out}}, e^{-(T_n - T_v)}) + J_v(\widetilde{B}_v^{(2)}(\delta_{\text{out}}, e^{-(T_n - T_v)}),\end{aligned}$$

and (3.4.4) becomes,

$$\begin{aligned}&\frac{1}{n}\mathbf{E}(N_{i,j}(n)) = \frac{1}{n}\sum_{v=1}^{n}\mathbf{P}\left[(I_v(T_n - T_v), O_v(T_n - T_v)) = (i, j)\right]\\ =&\frac{1}{n}\sum_{v=1}^{n}\mathbf{P}\left[\mathcal{B}_v^{(n)} = (i, j)\right] + \frac{1}{n}\left(\mathbf{P}\left[(I_1(T_n), O_1(T_n)) = (i, j)\right] - \mathbf{P}\left[\mathcal{B}_1^{(n)} = (i, j)\right]\right). \quad (3.4.6)\end{aligned}$$

The last step is necessitated by the construction since $(I_1(t), O_1(t))$ is a pair of independent BI processes, which is different from the rest of the $(I_v(\cdot), O_v(\cdot))_{v\geq 2}$ pairs. Here this difference is inconsequential because as $n \to \infty$,

$$\frac{1}{n}\left|\mathbf{P}\left[(I_1(T_n), O_1(T_n)) = (i, j)\right] - \mathbf{P}\left[\mathcal{B}_1^{(n)} = (i, j)\right]\right| \leq \frac{2}{n} \to 0.$$

So we only need to consider the first term in (3.4.6). Let U_n be a random variable uniformly distributed on $[n-1]$ and independent of the rest. Then

$$
\begin{aligned}
&\frac{1}{n}\sum_{v=1}^{n}\mathbf{P}\left[\mathcal{B}_v^{(n)}=(i,j)\right]\\
=&\alpha\frac{1}{n}\sum_{v=1}^{n}\mathbf{P}\left[(B_v^{(1)}(\delta_{\text{in}},e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(T_n-T_v)}),1+\widetilde{B}_v^{(1)}(1+\delta_{\text{out}},e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(T_n-T_v)}))=(i,j)\right]\\
&+\gamma\frac{1}{n}\sum_{v=1}^{n}\mathbf{P}\left[(1+B_v^{(2)}(1+\delta_{\text{in}},e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(T_n-T_v)}),\widetilde{B}_v^{(2)}(\delta_{\text{out}},e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(T_n-T_v)})=(i,j)\right]\\
=&\alpha\left(1-\frac{1}{n}\right)\mathbf{P}\left[(B_1^{(1)}(\delta_{\text{in}},e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(T_n-T_{U_n})}),1+\widetilde{B}_1^{(1)}(1+\delta_{\text{out}},e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(T_n-T_{U_n})}))=(i,j)\right]\\
&+\gamma\left(1-\frac{1}{n}\right)\mathbf{P}\left[(1+B_1^{(2)}(1+\delta_{\text{in}},e^{-\frac{c_{\text{in}}}{c_{\text{in}}+c_{\text{out}}}(T_n-T_{U_n})}),\widetilde{B}_1^{(2)}(\delta_{\text{out}},e^{-\frac{c_{\text{out}}}{c_{\text{in}}+c_{\text{out}}}(T_n-T_{U_n})})=(i,j)\right]\\
&+\frac{1}{n}\mathbf{P}\left[\mathcal{B}_n^{(n)}=(i,j)\right],
\end{aligned}
$$

since the distributions of $B_v^{(1)},\widetilde{B}_v^{(1)},B_v^{(2)},\widetilde{B}_v^{(1)}$ do not depend on v. Let T be a unit exponential random variable that is independent of I_v,O_v, $v\geq 1$. A variant of the Renyi representation for exponential order statistics (see [22, Theorem 3.14] for details) gives

$$T_n-T_{U_n}\overset{d}{=}\frac{T}{(c_{\text{in}}+c_{\text{out}})^{-1}}. \tag{3.4.7}$$

Define a Bernoulli random variable J that is independent from T, $B_1^{(1)}$, $B_1^{(2)}$, $\widetilde{B}_1^{(1)}$ and $\widetilde{B}_1^{(2)}$ with $\mathbf{P}(J=1)=\gamma=1-\mathbf{P}(J=0)$. Then applying (3.4.7) therefore gives

$$
\begin{aligned}
&\frac{1}{n}\sum_{v=1}^{n}\mathbf{P}\left[\mathcal{B}_v^{(n)}=(i,j)\right]\\
=&\alpha\left(1-\frac{1}{n}\right)\mathbf{P}\left[(B_1^{(1)}(\delta_{\text{in}},e^{-c_{\text{in}}T}),1+\widetilde{B}_1^{(1)}(1+\delta_{\text{out}},e^{-c_{\text{out}}T}))=(i,j)\right]\\
&+\gamma\left(1-\frac{1}{n}\right)\mathbf{P}\left[(1+B_1^{(2)}(1+\delta_{\text{in}},e^{-c_{\text{in}}T}),\widetilde{B}_1^{(2)}(\delta_{\text{out}},e^{-c_{\text{out}}T})=(i,j)\right]+\frac{1}{n}\left[\mathcal{B}_n^{(n)}=(i,j)\right]\\
=&\left(1-\frac{1}{n}\right)\mathbf{P}\Big[(1-J)(B_1^{(1)}(\delta_{\text{in}},e^{-c_{\text{in}}T}),1+\widetilde{B}_1^{(1)}(1+\delta_{\text{out}},e^{-c_{\text{out}}T}))\\
&\qquad+J(1+B_1^{(2)}(1+\delta_{\text{in}},e^{-c_{\text{in}}T}),\widetilde{B}_1^{(2)}(\delta_{\text{out}},e^{-c_{\text{out}}T})=(i,j)\Big]+\frac{1}{n}\left[\mathcal{B}_n^{(n)}=(i,j)\right]\\
=&\left(1-\frac{1}{n}\right)\mathbf{P}\left[(\mathcal{I},\mathcal{O})=(i,j)\right]+\frac{1}{n}\left[\mathcal{B}_n^{(n)}=(i,j)\right].
\end{aligned}
$$

Therefore,

$$\left|\frac{1}{n}\mathbf{E}\left[N_{ij}(n)\right] - \mathbf{P}\left[(\mathcal{I},\mathcal{O}) = (i,j)\right]\right| \leq \frac{4}{n},$$

which leads to (3.4.2) and (3.4.3) as $n \to \infty$. □

Remark 3.4.3. This argument also shows that for $x > 0, y > 0$,

$$\frac{1}{n}\mathbf{E}N_{>x,>y}(n) = \mathbf{P}((\mathcal{I},\mathcal{O}) \in (x,\infty] \times (y,\infty]) + \epsilon_n(x,y), \tag{3.4.8}$$

where

$$\sup_{x>0,y>0} |\epsilon_n(x,y)| \leq \frac{4}{n}.$$

3.4.2 Convergence of the joint empirical measure

In this section, we investigate the convergence of the joint empirical measure:

$$\frac{1}{k_n}\sum_{k=1}^{n} \epsilon_{\left(D_i^{\text{in}}(n)/b_1(n/k_n),\, D_i^{\text{out}}(n)/b_2(n/k_n)\right)}(\cdot),$$

with scaling functions $b_i(\cdot)$, $i = 1, 2$, and some intermediate sequence k_n such that $k_n/n \to 0$ and $k_n \to \infty$ as $n \to \infty$. From (3.4.1), we have

$$\frac{1}{n}\sum_{v\in[n]} \epsilon_{\left(D_v^{\text{in}}(n), D_v^{\text{out}}(n)\right)}(\{(i,j)\}) \xrightarrow{p} \mathbf{P}((\mathcal{I},\mathcal{O}) = (i,j)), \qquad n \to \infty. \tag{3.4.9}$$

Moreover, [55, Theorem 2] shows that the limit pair $(\mathcal{I},\mathcal{O})$ is non-standard regularly varying, i.e.

$$n\mathbf{P}\left[\left(\frac{\mathcal{I}}{n^{c_{\text{in}}}}, \frac{\mathcal{O}}{n^{c_{\text{out}}}}\right) \in \cdot\right] \xrightarrow{v} \gamma V_1(\cdot) + \alpha V_2(\cdot), \quad n \to \infty, \tag{3.4.10}$$

in $M_+([0,\infty]^2 \setminus \{\mathbf{0}\})$ and $V_i(\cdot)$, $i = 1, 2$, concentrate on $(0,\infty)^2$ with Lebesgue densities given below in (3.4.14) and (3.4.15). It is also shown in [67] that the density of the limit measure is jointly regularly varying, and the relationship between the regular variation of the limit measure and that of the limit density has been explained.

Let $b_1(t) = t^{c_{\rm in}}$ and $b_2(t) = t^{c_{\rm out}}$, then heuristically, combining (3.4.9) and (3.4.10) gives

$$\frac{1}{k_n}\sum_{v\in[n]}\epsilon_{\left(D_v^{\rm in}(n)/(n/k_n)^{c_{\rm in}},\,D_v^{\rm out}(n)/(n/k_n)^{c_{\rm out}}\right)}(\cdot) \approx \frac{n}{k_n}\mathbf{P}\left[\left(\frac{\mathcal{I}}{(n/k_n)^{c_{\rm in}}},\frac{\mathcal{O}}{(n/k_n)^{c_{\rm out}}}\right)\in\cdot\right] \tag{3.4.11}$$

$$\Rightarrow \gamma V_1(\cdot)+\alpha V_2(\cdot), \quad n\to\infty$$

in $\mathbb{M}([0,\infty]^2\setminus\{\mathbf{0}\})$. We justify the approximation in (3.4.11) and the convergence result is summarized in the following theorem.

Theorem 3.4.4. *Suppose that $\{k_n\}$ is an intermediate sequence satisfying*

$$\liminf_{n\to\infty} k_n/(n\log n)^{1/2}>0 \quad \textit{and} \quad k_n/n\to 0 \quad \textit{as} \quad n\to\infty, \tag{3.4.12}$$

and recall $a = c_{out}/c_{in}$. Then we have

$$\frac{1}{k_n}\sum_{v\in[n]}^{n}\epsilon_{\left(D_v^{in}(n)/(n/k_n)^{c_{in}},\,D_v^{out}(n)/(n/k_n)^{c_{out}}\right)}(\cdot)\Rightarrow \gamma V_1(\cdot)+\alpha V_2(\cdot), \tag{3.4.13}$$

in $M_+([0,\infty]^2\setminus\{\mathbf{0}\})$, where V_1 and V_2 concentrate on $(0,\infty)^2$ with Lebesgue densities

$$f_1(x,y)=\frac{x^{\delta_{in}}y^{\delta_{out}-1}}{c_{in}\Gamma(1+\delta_{in})\Gamma(\delta_{out})}\int_0^\infty z^{-(2+1/c_{in}+\delta_{in}+a\delta_{out})}e^{-x/z+y/z^a}\mathrm{d}z, \tag{3.4.14}$$

and

$$f_2(x,y)=\frac{x^{\delta_{in}-1}y^{\delta_{out}}}{c_{in}\Gamma(\delta_{in})\Gamma(1+\delta_{out})}\int_0^\infty z^{-(1+a+1/c_{in}+\delta_{in}+a\delta_{out})}e^{-x/z+y/z^a}\mathrm{d}z, \tag{3.4.15}$$

respectively.

Proof. Proving (3.4.13) requires using concentration results for degree counts $N_{i,j}(n)$ which compare counts with expected counts; these are collected in Chapter B.1. In this section we show for $x,y>0$,

$$\left|\frac{1}{k_n}\mathbf{E}\left(N_{>\left(\frac{n}{k_n}\right)^{c_{\rm in}}x,>\left(\frac{n}{k_n}\right)^{c_{\rm out}}y}(n)\right)-\frac{n}{k_n}\,p_{>\left(\frac{n}{k_n}\right)^{c_{\rm in}}x,>\left(\frac{n}{k_n}\right)^{c_{\rm out}}y}\right|\xrightarrow{p}0, \tag{3.4.16a}$$

$$\left|\frac{1}{k_n}\mathbf{E}\left(N^{\text{in}}_{>\left(\frac{n}{k_n}\right)^{c_{\text{in}}}x}(n)\right) - \frac{n}{k_n}\, p^{\text{in}}_{>\left(\frac{n}{k_n}\right)^{c_{\text{in}}}x}\right| \xrightarrow{p} 0, \tag{3.4.16b}$$

$$\left|\frac{1}{k_n}\mathbf{E}\left(N^{\text{out}}_{>\left(\frac{n}{k_n}\right)^{c_{\text{out}}}y}(n)\right) - \frac{n}{k_n}\, p^{\text{out}}_{>\left(\frac{n}{k_n}\right)^{c_{\text{out}}}y}\right| \xrightarrow{p} 0. \tag{3.4.16c}$$

We give a proof for (3.4.16a) and (3.4.16b) and (3.4.16c) follows from a similar argument.

Adopting the notation from the proof of Theorem 3.4.1, and using (3.4.8) we have

$$\begin{aligned}
&\left|\frac{1}{k_n}\mathbf{E}\left(N_{>\left(\frac{n}{k_n}\right)^{c_{\text{in}}}x,\,>\left(\frac{n}{k_n}\right)^{c_{\text{out}}}y}(n)\right) - \frac{n}{k_n}\, p_{>\left(\frac{n}{k_n}\right)^{c_{\text{in}}}x,\,>\left(\frac{n}{k_n}\right)^{c_{\text{out}}}y}\right| \\
&\quad= \left|\frac{n}{k_n}\frac{1}{n}\sum_{v\in[n]}\mathbf{P}\left(\frac{D^{\text{in}}_v(n)}{(n/k_n)^{c_{\text{in}}}} > x,\ \frac{D^{\text{out}}_v(n)}{(n/k_n)^{c_{\text{out}}}} > y\right) - \frac{n}{k_n}\mathbf{P}\left[\frac{\mathcal{I}}{(n/k_n)^{c_{\text{in}}}} > x,\ \frac{\mathcal{O}}{(n/k_n)^{c_{\text{out}}}} > y\right]\right| \\
&\quad= \left|\frac{n}{k_n}\frac{1}{n}\sum_{v=1}^{n}\mathbf{P}\left(\mathcal{B}^{(n)}_v \in \left((n/k_n)^{c_{\text{in}}}x,\infty\right]\times\left((n/k_n)^{c_{\text{out}}}y,\infty\right]\right)\right. \\
&\qquad\qquad \left. -\frac{n}{k_n}\mathbf{P}\left[\frac{\mathcal{I}}{(n/k_n)^{c_{\text{in}}}} > x,\ \frac{\mathcal{O}}{(n/k_n)^{c_{\text{out}}}} > y\right]\right| \\
&\qquad + \frac{1}{k_n}\left|\mathbf{P}\left(\frac{D^{\text{in}}_1(n)}{(n/k_n)^{c_{\text{in}}}} > x,\ \frac{D^{\text{out}}_1(n)}{(n/k_n)^{c_{\text{out}}}} > y\right)\right. \\
&\qquad\qquad\qquad \left. -\mathbf{P}\left(\mathcal{B}^{(n)}_1 \in \left((n/k_n)^{c_{\text{in}}}x,\infty\right]\times\left((n/k_n)^{c_{\text{out}}}y,\infty\right]\right)\right| \\
&\qquad \le \epsilon_n\left((n/k_n)^{c_{\text{in}}}x, (n/k_n)^{c_{\text{out}}}y\right) + \frac{2}{k_n} \to 0,
\end{aligned}$$

as $n \to \infty$.

Combining concentration results in (B.1.1), (B.1.5) and (B.1.6) with (3.4.16) implies that for any intermediate sequence $\{k_n\}$ satisfying (3.4.12) and $x, y > 0$, as $n \to \infty$,

$$\frac{1}{k_n}\left|N_{>\left(\frac{n}{k_n}\right)^{c_{\text{in}}}x,\,>\left(\frac{n}{k_n}\right)^{c_{\text{out}}}y}(n) - n\, p_{>\left(\frac{n}{k_n}\right)^{c_{\text{in}}}x,\,>\left(\frac{n}{k_n}\right)^{c_{\text{out}}}y}\right| \xrightarrow{p} 0, \tag{3.4.17a}$$

$$\frac{1}{k_n}\left|N^{\text{in}}_{>\left(\frac{n}{k_n}\right)^{c_{\text{in}}}x}(n) - n\, p^{\text{in}}_{>\left(\frac{n}{k_n}\right)^{c_{\text{in}}}x}\right| \xrightarrow{p} 0, \tag{3.4.17b}$$

$$\frac{1}{k_n}\left|N^{\text{out}}_{>\left(\frac{n}{k_n}\right)^{c_{\text{out}}}y}(n) - n\, p^{\text{out}}_{>\left(\frac{n}{k_n}\right)^{c_{\text{out}}}y}\right| \xrightarrow{p} 0. \tag{3.4.17c}$$

Define the vague metric $\rho(\cdot,\cdot)$ on $M_+([0,\infty]^2\setminus\{\mathbf{0}\})$ (cf. [50, Chapter 3.3]) as follows. There exists some sequence of continuous functions on $[0,\infty]^2\setminus\{\mathbf{0}\}$ with compact supports, $f_i : [0,\infty]^2 \setminus \{\mathbf{0}\} \mapsto \mathbb{R}_+$, $i \geq 1$, and for $\mu_1, \mu_2 \in M_+([0,\infty]^2 \setminus \{\mathbf{0}\})$,

$$\rho(\mu_1,\mu_2) = \sum_{i=1}^{\infty} \frac{|\mu_1(f_i) - \mu_2(f_i)| \min 1}{2^i},$$

where $\mu_j(f_i) := \int_{[0,\infty]^2\setminus\{\mathbf{0}\}} f_i(x)\mu_j(\mathrm{d}x)$, $j = 1,2$, $i \geq 1$. Then results in (3.4.17) imply: as $n \to \infty$,

$$\rho\left(\frac{1}{k_n}\sum_{v\in[n]}^{n} \epsilon_{\left(D_v^{\mathrm{in}}(n)/(n/k_n)^{c_{\mathrm{in}}},\, D_v^{\mathrm{out}}(n)/(n/k_n)^{c_{\mathrm{out}}}\right)}, \frac{n}{k_n}\mathbf{P}\left[\left(\frac{\mathcal{I}}{(n/k_n)^{c_{\mathrm{in}}}}, \frac{\mathcal{O}}{(n/k_n)^{c_{\mathrm{out}}}}\right) \in \cdot\right]\right) \xrightarrow{p} 0. \quad (3.4.18)$$

Then (3.4.13) follows from combining (3.4.18) and the vague convergence in (3.4.10), with (3.4.14) and (3.4.15) being specified in [55, Theorem 2]. ◻

3.5 Consistency of the Hill Estimator

In practice, the growth rates of in- and out-degrees are often estimated by Hill estimators as defined in (3.1.4). However, despite its wide use, there is no theoretical justification for such estimates and the consistency has been proved only for a simple undirected preferential attachment model in [68]. We now turn to (3.1.7) and (3.1.8) as preparations for considering consistency of the Hill estimator.

Proposition 3.5.1. *Suppose that $\{k_n\}$ is some intermediate sequence satisfying* (3.4.12)*. Define*

$$b_1(t) = \left[c_{in}\frac{\Gamma(1+\delta_{in}+c_{in}^{-1})}{\Gamma(1+\delta_{in})}\left(\frac{\alpha\delta_{in}}{1+c_{in}\delta_{in}} + \frac{\gamma}{c_{in}}\right)\right]^{c_{in}} t^{c_{in}},$$
$$b_2(t) = \left[c_{out}\frac{\Gamma(1+\delta_{out}+c_{out}^{-1})}{\Gamma(1+\delta_{out})}\left(\frac{\gamma\delta_{out}}{1+c_{out}\delta_{out}} + \frac{\alpha}{c_{out}}\right)\right]^{c_{out}} t^{c_{out}},$$

then

$$\frac{1}{k_n}\sum_{v\in[n]}\epsilon_{D_v^{in}(n)/b_1(n/k_n)} \Rightarrow \nu_{c_{in}^{-1}}, \qquad \textit{in } M_+((0,\infty]), \tag{3.5.1}$$

$$\frac{1}{k_n}\sum_{v\in[n]}\epsilon_{D_v^{out}(n)/b_2(n/k_n)} \Rightarrow \nu_{c_{out}^{-1}}, \qquad \textit{in } M_+((0,\infty]). \tag{3.5.2}$$

Proof. Marginalizing the results in (3.4.13) gives

$$\frac{1}{k_n}\sum_{v\in[n]}\epsilon_{\frac{D_v^{\text{in}}(n)}{(n/k_n)^{c_{\text{in}}}}} \Rightarrow c_{\text{in}}\frac{\Gamma(1+\delta_{\text{in}}+c_{\text{in}}^{-1})}{\Gamma(1+\delta_{\text{in}})}\left(\frac{\alpha\delta_{\text{in}}}{1+c_{\text{in}}\delta_{\text{in}}}+\frac{\gamma}{c_{\text{in}}}\right)\nu_{c_{\text{in}}^{-1}}, \qquad \text{in } M_+((0,\infty]),$$

$$\frac{1}{k_n}\sum_{v\in[n]}\epsilon_{\frac{D_v^{\text{out}}(n)}{(n/k_n)^{c_{\text{out}}}}} \Rightarrow c_{\text{out}}\frac{\Gamma(1+\delta_{\text{out}}+c_{\text{out}}^{-1})}{\Gamma(1+\delta_{\text{out}})}\left(\frac{\gamma\delta_{\text{out}}}{1+c_{\text{out}}\delta_{\text{out}}}+\frac{\alpha}{c_{\text{out}}}\right)\nu_{c_{\text{out}}^{-1}}, \qquad \text{in } M_+((0,\infty]).$$

Scaling both sides by the constant appearing in the limit measure gives (3.5.1) and (3.5.2). □

With Proposition 3.5.1 available, we now prove the consistency of Hill estimators for in- and out-degrees.

Theorem 3.5.2. *Let*

$$D_{(1)}^{in}(n) \ge D_{(2)}^{in}(n) \ge \cdots \ge D_{(n)}^{in}(n),$$

$$D_{(1)}^{out}(n) \ge D_{(2)}^{out}(n) \ge \cdots \ge D_{(n)}^{out}(n),$$

be order statistics for in- and out-degrees $\{D_v^{in}(n)\}_{v\in[n]}$, $\{D_v^{out}(n)\}_{v\in[n]}$, respectively. Define the Hill estimators for $\{D_v^{in}(n)\}_{v\in[n]}$ and $\{D_v^{out}(n)\}_{v\in[n]}$ as

$$H_{k,n}^{in} := \frac{1}{k}\sum_{i=1}^{k}\log\frac{D_{(i)}^{in}(n)}{D_{(k+1)}^{in}(n)}, \qquad H_{k,n}^{out} := \frac{1}{k}\sum_{i=1}^{k}\log\frac{D_{(i)}^{out}(n)}{D_{(k+1)}^{out}(n)}.$$

Then for some intermediate sequence $\{k_n\}$ satisfying (3.4.12)*, we have as $n\to\infty$,*

$$H_{k_n,n}^{in} \xrightarrow{p} c_{in}, \qquad \textit{and} \qquad H_{k_n,n}^{out} \xrightarrow{p} c_{out}. \tag{3.5.3}$$

Proof. From (3.5.1) and (3.5.2), we conclude by inversion and [50, Proposition 3.2] that in $D(0,\infty]$

$$\frac{D^{\text{in}}_{([k_n t])}(n)}{b_1(n/k_n)} \xrightarrow{p} t^{-c_{\text{in}}} \quad \text{and} \quad \frac{D^{\text{out}}_{([k_n t])}(n)}{b_2(n/k_n)} \xrightarrow{p} t^{-c_{\text{out}}}.$$

Therefore,

$$\left(\frac{1}{k_n}\sum_{v\in[n]} \epsilon_{D^{\text{in}}_v(n)/b_1(n/k_n)}, \frac{D^{\text{in}}_{(k_n)}(n)}{b_1(n/k_n)}\right) \Rightarrow (\nu_{c_{\text{in}}^{-1}}, 1) \quad \text{in } M_+((0,\infty])\times(0,\infty), \tag{3.5.4}$$

$$\left(\frac{1}{k_n}\sum_{v\in[n]} \epsilon_{D^{\text{out}}_v(n)/b_2(n/k_n)}, \frac{D^{\text{out}}_{(k_n)}(n)}{b_2(n/k_n)}\right) \Rightarrow (\nu_{c_{\text{out}}^{-1}}, 1) \quad \text{in } M_+((0,\infty])\times(0,\infty). \tag{3.5.5}$$

Define the operator

$$S : M_+((0,\infty])\times(0,\infty) \mapsto M_+((0,\infty])$$

by

$$S(\nu, c)(A) = \nu(cA).$$

By the proof in [50, Theorem 4.2], the mapping S is continuous at $(\nu_{c_i^{-1}}, 1)$, $i = 1, 2$. Therefore, applying the continuous mapping S to the joint weak convergence in (3.5.4) and (3.5.5) gives

$$\frac{1}{k_n}\sum_{v\in[n]} \epsilon_{D^{\text{in}}_v(n)\big/D^{\text{in}}_{(k_n)}(n)} \Rightarrow \nu_{c_{\text{in}}^{-1}}, \qquad \text{in } M_+((0,\infty]),$$

$$\frac{1}{k_n}\sum_{v\in[n]} \epsilon_{D^{\text{out}}_v(n)\big/D^{\text{out}}_{(k_n)}(n)} \Rightarrow \nu_{c_{\text{out}}^{-1}}, \qquad \text{in } M_+((0,\infty]).$$

Then the rest of the proof is similar to arguments in the proof of [68, Theorem 11]. Here we only include proofs for the consistency $H^{\text{in}}_{k_n,n}$ and that for $H^{\text{out}}_{k_n,n}$ follows from the same argument. Define $\hat{\nu}^{\text{in}}_n(\cdot) := \frac{1}{k_n}\sum_{v\in[n]} \epsilon_{D^{\text{in}}_v(n)\big/D^{\text{in}}_{(k_n)}(n)}(\cdot)$. First observe

$$H^{\text{in}}_{k_n,n} = \int_1^\infty \hat{\nu}^{\text{in}}_n(y,\infty]\frac{\mathrm{d}y}{y}.$$

Then fix $M > 0$ large and define a mapping $f \mapsto \int_1^M f(y)\frac{\mathrm{d}y}{y}$ from $D(0,\infty] \mapsto \mathbb{R}_+$. This map is a.s. continuous so

$$\int_1^M \hat{\nu}_n^{\text{in}}(y,\infty]\frac{\mathrm{d}y}{y} \xrightarrow{p} \int_1^M \nu_{c_{\text{in}}^{-1}}(y,\infty]\frac{\mathrm{d}y}{y},$$

and it remains to show by the second converging together theorem (cf. [50, Theorem 3.5]) that

$$\lim_{M\to\infty}\limsup_{n\to\infty} \mathbf{P}\left(\int_M^\infty \hat{\nu}_n^{\text{in}}(y,\infty]\frac{\mathrm{d}y}{y} > \varepsilon\right) = 0. \tag{3.5.6}$$

The probability in (3.5.6) is

$$\begin{aligned}
\mathbf{P}\left(\int_M^\infty \hat{\nu}_n^{\text{in}}(y,\infty]\frac{\mathrm{d}y}{y} > \varepsilon\right) \le\ & \mathbf{P}\left(\int_M^\infty \hat{\nu}_n^{\text{in}}(y,\infty]\frac{\mathrm{d}y}{y} > \varepsilon, \left|\frac{D_{(k_n)}^{\text{in}}(n)}{b_1(n/k_n)} - 1\right| < \eta\right)\\
&+ \mathbf{P}\left(\int_M^\infty \hat{\nu}_n^{\text{in}}(y,\infty]\frac{\mathrm{d}y}{y} > \varepsilon, \left|\frac{D_{(k_n)}^{\text{in}}(n)}{b_1(n/k_n)} - 1\right| \ge \eta\right)\\
\le\ & \mathbf{P}\left(\int_M^\infty \frac{1}{k_n}\sum_{i=1}^n \epsilon_{D_i^{\text{in}}(n)/b_1(n/k_n)}((1-\eta)y,\infty]\frac{\mathrm{d}y}{y} > \varepsilon\right)\\
&+\mathbf{P}\left(\left|\frac{D_{(k_n)}^{\text{in}}(n)}{b_1(n/k_n)} - 1\right| \ge \eta\right) =: A + B.
\end{aligned}$$

By (3.5.4), $B \to 0$ as $n \to \infty$, and using the Markov inequality, A is bounded by

$$\begin{aligned}
&\frac{1}{\varepsilon}\mathbf{E}\left(\int_M^\infty \frac{1}{k_n}\sum_{v=1}^n \epsilon_{D_v^{\text{in}}(n)/b_1(n/k_n)}((1-\eta)y,\infty]\frac{\mathrm{d}y}{y}\right)\\
=&\frac{1}{\varepsilon}\mathbf{E}\left(\int_{M(1-\eta)}^\infty \frac{1}{k_n}\sum_{v=1}^n \epsilon_{D_v^{\text{in}}(n)/b_1(n/k_n)}(y,\infty]\frac{\mathrm{d}y}{y}\right) \le \frac{1}{\varepsilon}\int_{M(1-\eta)}^\infty \frac{1}{k_n}\mathbf{E}\left(N_{>[b_1(n/k_n)y]}^{\text{in}}(n)\right)\frac{\mathrm{d}y}{y}.
\end{aligned}$$

Using Stirling's formula, (3.4.17b) gives that for $y > 0$,

$$\frac{1}{k_n}\mathbf{E}\left(N_{>[b_1(n/k_n)y]}^{\text{in}}(n)\right) \to y^{-c_{\text{in}}^{-1}}. \tag{3.5.7}$$

Let $U(t) := \mathbf{E}(N_{>t}^{\text{in}}(n))$ and (3.5.7) becomes: for $y > 0$,

$$\frac{1}{k_n}U(b_1(n/k_n)y) \to y^{-c_{\text{in}}^{-1}}, \quad \text{as } n \to \infty.$$

Since $U(\cdot)$ is a non-increasing function, $U \in RV_{-c_{\text{in}}^{-1}}$ by [50, Proposition 2.3(ii)]. Therefore, Karamata's theorem gives

$$A \leq \frac{1}{\varepsilon} \int_{M(1-\eta)}^{\infty} \frac{1}{k_n} \mathbf{E}\left(N^{\text{in}}_{>[b_1(n/k_n)y]}(n)\right) \frac{\mathrm{d}y}{y} \sim C(\delta, \eta) M^{-c_{\text{in}}^{-1}},$$

with some positive constant $C(\delta, \eta) > 0$. Also, $M^{-c_{\text{in}}^{-1}} \to 0$ as $M \to \infty$, and (3.5.6) follows. $\square$

CHAPTER 4

THRESHOLD SELECTION

4.1 Overview

Power-laws are ubiquitous in social network modeling and power law index estimation reveals important characteristics. In particular, it allows to conclude the likelihood of having a node with large degrees. Data repositories of large network datasets such as KONECT [42] (`http://konect.cc/`) provide estimates of power-law tail indices as one of the key summary statistics for almost all listed networks, and these estimates are obtained by computing the *Hill estimate* [28] of the power-law index of the degree distribution. A brief summary of Hill estimation is given in Chapter 4.1.1. Figure 4.1.1 presents part of the statistical summaries for the Flickr friendship data given on KONECT [42] (`http://konect.cc/networks/flickrEdges/`), and "power law exponent" corresponds to the power-law index Hill estimate of the degree frequencies.

When estimating the power-law index, an important step is to determine the cutoff value so that the distribution of observations larger than this thresh-

Square count	428,143,604,855
4-tour count	3,483,825,597,342
Power law exponent (estimated) with d_{min}	1.7310 (d_{min} = 8)
Gini coefficient	88.2%
Relative edge distribution entropy	82.2%
Assortativity	-0.015282

Figure 4.1.1: A snapshot of summary statistics for the Flickr friendship data. The full list of key statistics is available at http://konect.uni-koblenz.de/networks/flickr-links.

old follows a power law. Following the notation used in KONECT, we need to determine the value of $d_{\min}$, the minimal degree, above which the degree distribution follows a power law. A threshold selection procedure is proposed in [10], where one chooses the cutoff value that yields the smallest Kolmogorov-Smirnov distance between the empirical distribution above the threshold and the corresponding fitted power law. This selection procedure is for instance widely adopted in the analyses of social networks and income distributions, having attracted more than 3,000 citations; see, for example, [1, 9, 30, 34, 43, 47, 54, 56]. It has also been encoded as an R-package called `poweRlaw` (cf. [24]). We now outline this threshold selection method.

4.1.1 Minimum distance selection procedure (MDSP)

Mathematically, a nonnegative random variable X follows a power law distribution if its tail distribution function satisfies

$$1 - F(x) = cx^{-\alpha}, \tag{4.1.1}$$

for x exceeding some threshold $x_0 > 0$, where $c > 0$ is some constant and $\alpha > 0$ is known as the exponent or tail index. The distribution that fulfills this relation for all $x > c^{1/\alpha}$ is called Pareto distribution. Therefore, it is also common to speak of a Pareto tail instead of a power tail.

As mentioned in [10], empirical distributions rarely follow a power law for all values, but rather only for observations greater than some cutoff value. Therefore, there are two parameters to determine: the exponent α and the cutoff value x_0. Provided that we have a good estimate for the threshold x_0, we can discard all observations below x_0 and estimate α by the maximum likelihood estimator based on the remaining exceedances. More precisely, suppose random variables X_i, $1 \leq i \leq n$, are observed and recall the notations in (3.1.4). If one uses the k-th largest order statistic as a threshold, the maximum likelihood approach applied to model (4.1.1) leads to the well-known Hill estimator [28]

$$\hat{\alpha}_{n,k} := \left(\frac{1}{k-1} \sum_{i=1}^{k-1} \log \frac{X_{(i)}}{X_{(k)}} \right)^{-1}. \tag{4.1.2}$$

This estimator is known to be consistent [45] and asymptotically normal [26] with rate $k^{-1/2}$, provided $X_{(k)}$ exceeds the threshold with probability tending to 1. Hence its performance strongly depends on the appropriate choice of k, the selection of which in turn requires an accurate estimate of x_0.

Clauset and Newman [10] suggest estimating the threshold by the order statistic which minimizes the Kolmogorov-Smirnov distance between the empirical distribution of the exceedances and the Pareto distribution fitted with the larger order statistics. To be more precise, define the Kolmogorov-Smirnov distance

$$D_k := \sup_{y \geq 1} \left| \frac{1}{k-1} \sum_{i=1}^{n} 1_{(y,\infty)}\Big(\frac{X_{(i)}}{X_{(k)}}\Big) - y^{-\hat{\alpha}_{n,k}} \right| \tag{4.1.3}$$

and use $X_{(k_n^*)}$ with

$$k_n^* := \operatorname*{argmin}_{k \in \{2,\dots,n\}} D_k \tag{4.1.4}$$

as an estimator of the unknown threshold. (If the point of minimum is not unique, we may e.g. choose the one with the smallest index.) Since we choose the threshold that minimizes the distance between fitted and empirical tail this method is called the minimum distance selection procedure (MDSP). This method has also been adapted to binned data in [62].

The method is widely applied in practice, particularly in Computer Science and Network Science. However, to the best of our knowledge its performance has not been mathematically analyzed even in classical contexts where data is assumed to come from an iid model of repeated sampling. We will show that the MDSP often leads to choosing a k_n^* that is too small, resulting in increased variance and root mean squared error (RMSE) for the Hill estimator relative to a choice which minimizes the asymptotic mean squared error.

We begin in Chapter 4.2 with the iid case assuming the underlying distribution is exact Pareto and thus $k = n$ would be the best choice for minimizing asymptotic RMSE. It will be shown that the distribution of k_n^*/n can be approximated by a distribution supported by the whole interval $(0, 1]$, so that with non-negligible probability k_n^* is much smaller than n. In Chapter 4.3, we discuss numerical results for the performance of the MDSP applied to the in-degrees of linear preferential attachment networks. All proofs are postponed to the appendix.

The MDSP offers attractive features. The procedure yields estimates without requiring user discretion. It is readily implemented with R-packages that are well designed and can be ported into another algorithm. In network simulations the tail index estimates provided by MDSP have often appeared reasonable, provided network parameters are close to those observed in empirical studies, cf. Chapter 4.3.1. However, this method has limitations and needs be applied with caution. Even in the classical iid case, the asymptotic theory of MDSP estimation is fairly complex and it is not an easy task to extract confidence intervals for estimates chosen with this method. Furthermore MDSP estimates of the tail index do not achieve minimal asymptotic RMSE. For the node based data of random graphs, there is no theoretical analysis available for MDSP estimates.

4.2 The Pareto case

Throughout this section, we assume that the observations are independently drawn from an exact Pareto distribution, that is

$$1 - F(x) = cx^{-\alpha}, \quad x \geq c^{1/\alpha}, \tag{4.2.1}$$

for some $\alpha, c > 0$. Such a model rarely arises in practice, but we will see that one of the main drawbacks of the MDSP can be most easily explained in this setting. In the case of (4.2.1), a reasonable selection procedure should pick some value k close to n because the whole survival function is a power function. In fact, since the choice $k = n$ give the MLE, it minimizes the mean squared error of the Hill estimator.

Note that

$$\begin{aligned} D_k &= \sup_{y \geq 1} \left| \frac{1}{k-1} \sum_{i=1}^{k-1} 1_{(y,\infty)}\Big(\frac{X_{(i)}}{X_{(k)}}\Big) - y^{-\hat{\alpha}_{n,k}} \right| \\ &= \max_{1 \leq j < k} \max \left(\Big(\frac{X_{(j)}}{X_{(k)}}\Big)^{-\hat{\alpha}_{n,k}} - \frac{j-1}{k-1}, \frac{j}{k-1} - \Big(\frac{X_{(j)}}{X_{(k)}}\Big)^{-\hat{\alpha}_{n,k}} \right) \\ &= \max_{1 \leq j \leq k} \left| \Big(\frac{X_{(j)}}{X_{(k)}}\Big)^{-\hat{\alpha}_{n,k}} - \frac{j}{k} \right| + rem \end{aligned} \tag{4.2.2}$$

with *rem* denoting a remainder term with modulus of at most $1/k$. It is well known that $n^{1/2} D_n$ weakly converges to the supremum of a Brownian bridge if the Hill estimator is replaced with the true value α. More generally, Theorem 4.2.1 given below shows that $n^{1/2} D_{\lceil nt \rceil}$ converges to $\sup_{s \in (0,1]} |Z(s,t)|$ for some Gaussian process uniformly for all $t \in [\varepsilon, 1]$ for any $\varepsilon \in (0,1)$. The limit process is self-similar with $\sup_{s \in (0,1]} |Z(s,t)| =^d t^{-1/2} \sup_{s \in (0,1]} |Z(s,1)|$. Hence, it is more likely that its infimum is attained at some t close to 1 than in the neighborhood of some smaller value. However, with non-negligible probability the point of minimum of $t \mapsto \sup_{s\in(0,1]} |Z(s,t)|$ is considerably smaller than 1, corresponding to a suboptimal behavior of the MDSP.

Theorem 4.2.1. *If $\{X_i : 1 \le i \le n\}$, are iid with cdf F given in* (4.2.1)*, then there exists a sequence of Brownian motions W_n such that (for suitable versions of X_i)*

$$n^{1/2} D_{\lceil nt \rceil} = \sup_{0<s\le 1} |Z_n(s,t)| + O_P\Big(\frac{\log(nt)(\log(nt) + (\log n)^{1/2})}{n^{1/2} t}\Big) \tag{4.2.3}$$

uniformly for $t \in [2/n, 1]$ with

$$Z_n(s,t) := \Big(\frac{W_n(st)}{st} - \frac{W_n(t)}{t}\Big)s + \Big(\int_0^1 \frac{W_n(tx)}{tx}\,dx - \frac{W_n(t)}{t}\Big)s \log s.$$

From Theorem 4.2.1 we obtain the joint asymptotic distribution of the selected number k and the resulting Hill estimator provided the limiting process has a unique point of minimum:

Corollary 4.2.2. *If $t \mapsto \sup_{0<s\le 1} |Z_1(s,t)|$ has a unique point of minimum T a.s., then*

$$(k_n^*/n, n^{1/2}(\hat\alpha_{n,k_n^*} - \alpha)) \Rightarrow \Big(T, \alpha\Big(\int_0^1 \frac{W_1(Tx)}{Tx}\,dx - \frac{W_1(T)}{T}\Big)\Big). \tag{4.2.4}$$

Remark 4.2.3. Unfortunately, the standard techniques to prove the uniqueness of the point of minimum of a Gaussian process apparently do not carry over to our limit process. However, the simulations outlined below suggest that indeed (4.2.4) holds. In any case, Theorem 3 of [21] implies the following weaker result:

If all points of minimum of $t \mapsto \sup_{0<s\le 1} |Z_1(s,t)|$ lie in a random interval $[T_0, T_1]$, then

$$P\{T_1 < x\} \le \liminf_{n\to\infty} P\{\hat k_n^*/n < x\} \le \limsup_{n\to\infty} P\{\hat k_n^*/n \le x\} \le P\{T_0 \le x\}$$

for all $x \in [0,1]$.

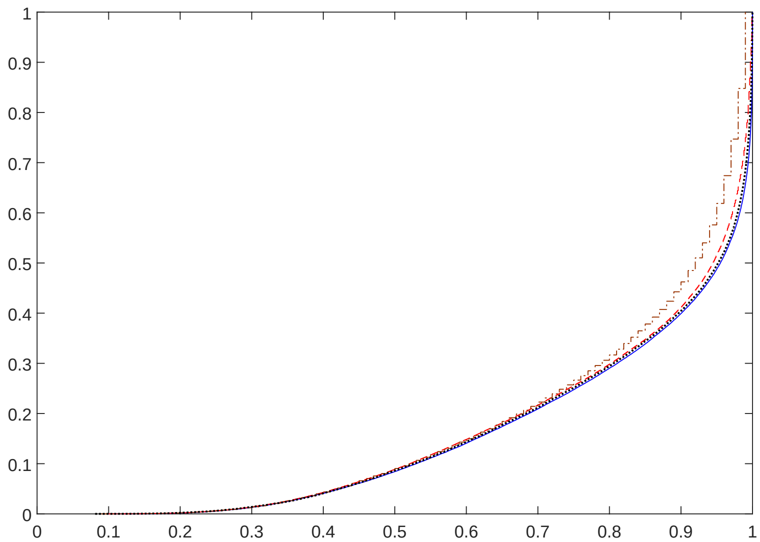


Figure 4.2.1: Empirical cdf of k_n^*/n for $n = 100$ (brown, dash-dotted), $n = 1,000$ (red, dashed) and $n = 10,000$ (black, dotted) and limit cdf according to (4.2.4) (blue, solid) for a Pareto model with $\alpha = c = 1$

Figure 4.2.1 shows the empirical cdf of k_n^*/n calculated from 10^5 simulations of standard Pareto samples (i.e., $\alpha = c = 1$) of size $n \in \{100; 1,000; 10,000\}$ in comparison with the limit cdf from (4.2.4). (The latter was approximately calculated from 10^5 simulations of a discretized version of the limit process Z_1 on a grid with $5 \cdot 10^4$ points in each argument.) The difference between the cdf of k_n^*/n and the limit cdf of T is small for $n = 1,000$ and hardly visible for $n = 10,000$, while k_n^*/n is stochastically a bit smaller for $n = 100$.

Reading from Figure 4.2.1, in the limit, the probability that k_n^* is less than $(3/4)n$ is about $1/4$, and the corresponding probabilities for $n/2$ and $n/3$ are 8.4% and 2%, respectively. So while in about $3/5$ of all cases k_n^* is at least $0.9n$, there is a non-negligible probability that k_n^* is substantially smaller than n. As a consequence, the variance and the root mean squared error (RMSE) of the corresponding Hill estimator $\hat{\alpha}_{n,k_n^*}$ are much larger than those of the estimator $\hat{\alpha}_{n,n}$

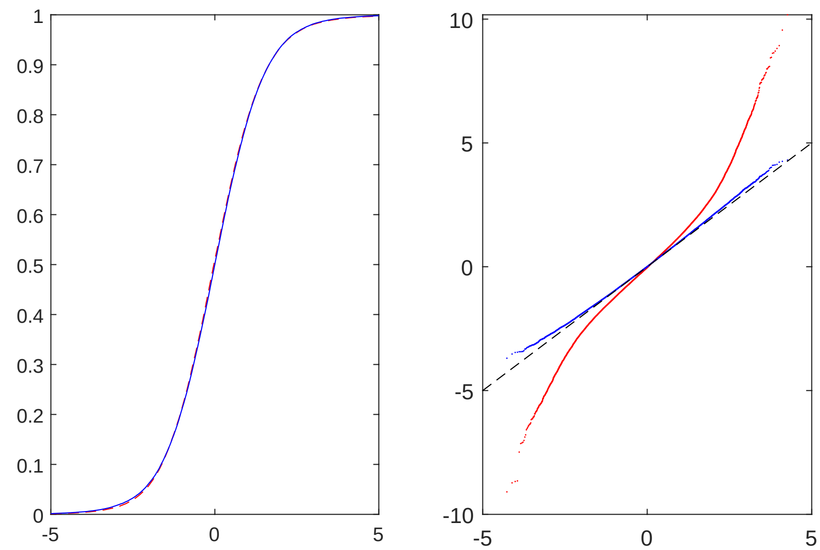


Figure 4.2.2: Left: empirical cdf of $n^{1/2}(\hat{\alpha}_{n,k_n^*} - \alpha)$ for $n = 1,000$ (red, dashed) and limit cdf according to (4.2.4) (blue, solid); right: normal Q-Q plot of $n^{1/2}(\hat{\alpha}_{n,k_n^*} - \alpha)$ (red) and of $n^{1/2}(\hat{\alpha}_{n,n} - \alpha)$ (blue) for $n = 1,000$, the black dashed line is the main diagonal

with minimal RMSE. In the limit, the variance of $\hat{\alpha}_{n,k_n^*}$ is about 88% larger, resulting in an RMSE which is 37% higher. For finite sample sizes the corresponding figures are 33% for $n = 100$, 36% for $n = 1,000$ and 37% for $n = 10,000$.

The left plot of Figure 4.2.2 compares the (empirical) distribution of $n^{1/2}(\hat{\alpha}_{n,k_n^*} - \alpha)$ for $n = 1,000$ with the limit distribution given in (4.2.4). Here the approximation is even better than in Figure 4.2.1. The right plot shows a normal Q-Q plot of the standardized estimation errors of $\hat{\alpha}_{n,k_n^*}$ and $\hat{\alpha}_{n,n}$, respectively. While the latter estimator is asymptotically normal (as is $\hat{\alpha}_{n,\lceil nt \rceil}$ for all $t \in (0,1]$), the Hill estimator based on the top k_n^* order statistics has much heavier tails.

The heavier tail of $\hat{\alpha}_{n,k_n^*}$ is important when constructing confidence intervals. The analyses in [10] tempt one to use $I_{1-\beta} := [\hat{\alpha}_{n,k_n^*}(1 - (k_n^*)^{-1/2} c_{\beta/2}); \hat{\alpha}_{n,k_n^*}(1 +$

$(k_n^*)^{-1/2}c_{\beta/2})]$, with $c_{\beta/2} = \Phi^{-1}(1-\beta/2)$ being the standard normal quantile to the level $1-\beta/2$, as the confidence interval with asymptotic level $1-\beta$. This approach, however, ignores the inherently stochastic nature of k_n^* and leads to a severe underestimation of the actual error. For example, the non-coverage probability of $I_{0.95}$ is greater than 8% and the one of $I_{0.99}$ larger than 2% for $n \in \{100; 1,000\}$. In contrast, the confidence interval derived from (4.2.4)

$$I^*_{1-\beta} := [\hat{\alpha}_{n,k_n^*}(1 - n^{-1/2}c^*_{\beta/2}); \hat{\alpha}_{n,k_n^*}(1 + n^{-1/2}c^*_{\beta/2})]$$

with $c^*_{\beta/2}$ denoting the $(1-\beta/2)$-quantile of the limit distribution of the standardized estimation error is very accurate. For $\beta = 5\%$ and $c^*_{\beta/2} \approx 2.74$ the non-coverage probability is 4.8%, for $\beta = 1\%$ and $c^*_{\beta/2} \approx 4.09$ it equals 1%.

Remark 4.2.4. In extreme value theory, it is often not assumed that above some threshold the tail is exactly of Pareto type, but that the difference between the actual tail and the approximating Pareto vanishes as the threshold increases. More precisely, a so-called second order condition may be used, e.g., that $F^{\leftarrow}(1-tx)/F^{\leftarrow}(1-t) - x^{1/\alpha} \sim t^{\rho}g(x)$ as $t \downarrow 0$ for some non-degenerate function g and some $\rho > 0$. In such a setting, a plethora of methods for selecting k aiming at a minimal RMSE of the Hill estimator have been suggested; see, for instance, [4], Chapter 4.7, or [25] for a comparison of some of these procedures.

Using the approach employed in the proof of Theorem 4.2.1, one can analyze the behavior of the MDSP in such a framework, too, provided the minimum of the Kolmogorov-Smirnov distance D_k is considered only over a set of indices $k \in \{2, \dots, k_n\}$ for some so-called intermediate sequence k_n, i.e., $k_n \to \infty$, but $k_n/n \to 0$. (A related result can be found in [38] which considers the Kolmogorov-Smirnov distance for an intermediate sequence converging to ∞ sufficiently slowly such that the deviation from the Pareto model is asymptotically negligible.) It turns out that this procedure is not able to pick a value k that asymptotically minimizes the RMSE of the Hill estimator. Moreover, simulations show that in terms of the RMSE of the Hill estimator, it is usually outperformed by other methods like the sequential procedure proposed by [18] or the bootstrap approach examined by [13].

4.3 Linear preferential attachment (PA) networks

One important application of the MDSP is the network models, where the power-law behavior of the degree distribution is widely observed. Theoretically, the linear PA model asymptotically generates power-law degree distributions and is therefore a popular choice to model networks. In this section, we first give an overview of the linear PA model and discuss the tail behavior of the in- and out-degrees, and then summarize simulation results on the performance of the MDSP.

Two parametric estimation methods for this directed linear PA model are derived in Chapters 2.2 and 2.3, giving estimates of ι_{in} and ι_{out} by simply plugging in the estimated parameters into (1.1.5) and (1.1.6), respectively. However, these estimates rely heavily on the correctness of the underlying model, which is hard to guarantee for real data. In Chapter 2.6, we propose another estimation method by coupling the Hill estimation of marginal degree distribution tail indices with the MDSP. Despite the dependence structure of degree sequences, the validity of the Hill estimation for linear PA models has been addressed in [65, 68], and numerical comparisons in [63] indicate that MDSP leads to more robust estimates against modeling error and data corruption.

4.3.1 Simulations

We now further examine the performance of the MDSP in the context of linear PA models through simulations. We simulate 10,000 linear PA graphs with expected number $m = 10^6$ of edges. To this end, starting from a trivial core with just one node and no edges, we grow the network as described above until it has $n = \lceil(\alpha + \gamma)m\rceil$ nodes. Then we try to find an appropriate number k such that the distribution of the k largest observed in-degrees can be well fitted by a power tail.

Here we report the results for two examples of parameters estimated from the KONECT [42] data sets:

I. *Baidu related pages*
(`http://konect.cc/networks/zhishi-baidu-relatedpages/`);

II. *Facebook wall posts*
(`http://konect.cc/networks/facebook-wosn-wall/`).

Using the snapshot methodology described in Chapter 2.3 we estimate

Example I: $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}}) = (0.0978, 0.873, 0.0289, 2.05, 13.8)$ resulting in $\iota_{\text{in}} = 1.30$;

Example II: $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}}) = (0.0327, 0.946, 0.0209, 8.88, 9.59)$ which implies $\iota_{\text{in}} = 1.51$.

(The parameters are rounded to 3 significant digits.)

In Example I the RMSE of the Hill estimator $\hat{\alpha}_{n,k_n^*}$ for the tail index of the in-degree is just 6.8% larger than the minimal RMSE over all deterministic choices

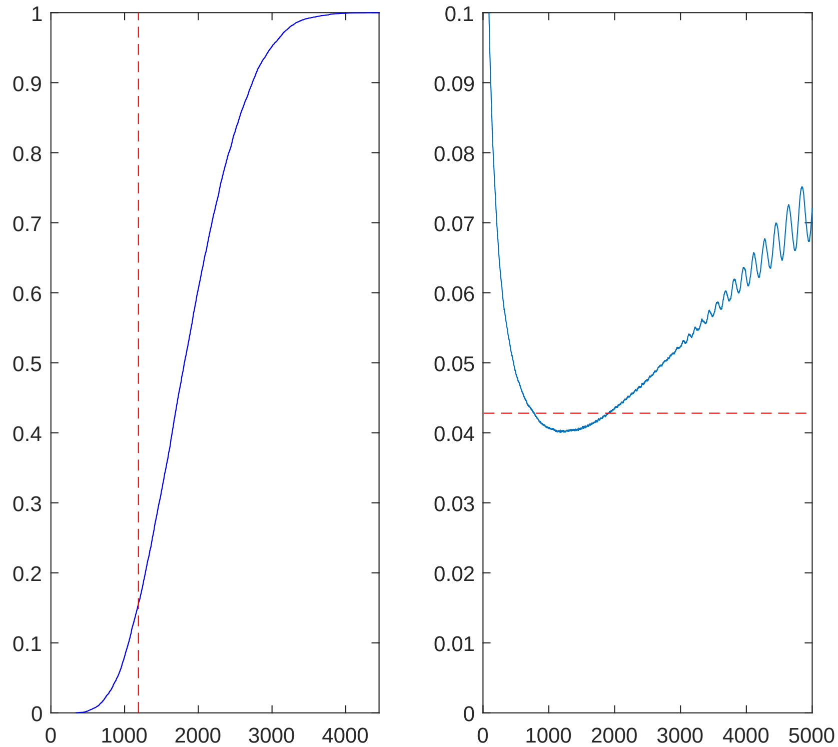


Figure 4.3.1: Left: empirical cdf of k_n^* for the linear PA Model I, the RMSE minimizing value of k is indicated by the dashed red line; right: RMSE of the Hill estimator vs. k, the RMSE of $\hat{\alpha}_{n,k_n^*}$ is indicated by the dashed red line.

of $k \in \{10, \dots, 10000\}$ which is attained for $k = 1187$. In this respect, the MDSP works much better for this linear PA network model than in any situation considered in Chapter 4.2. The right plot of Figure 4.3.1, which shows the RMSE of the Hill estimator as a function of k, hints at the reason for this good performance. There is a wide range of values k that lead to almost the same RMSE. So although the distribution of k_n^* (shown in the left plot) is spread out over the interval $[500, 3500]$, this does not increase the RMSE substantially.

In the Facebook Example II, the loss of efficiency is much larger. Here the

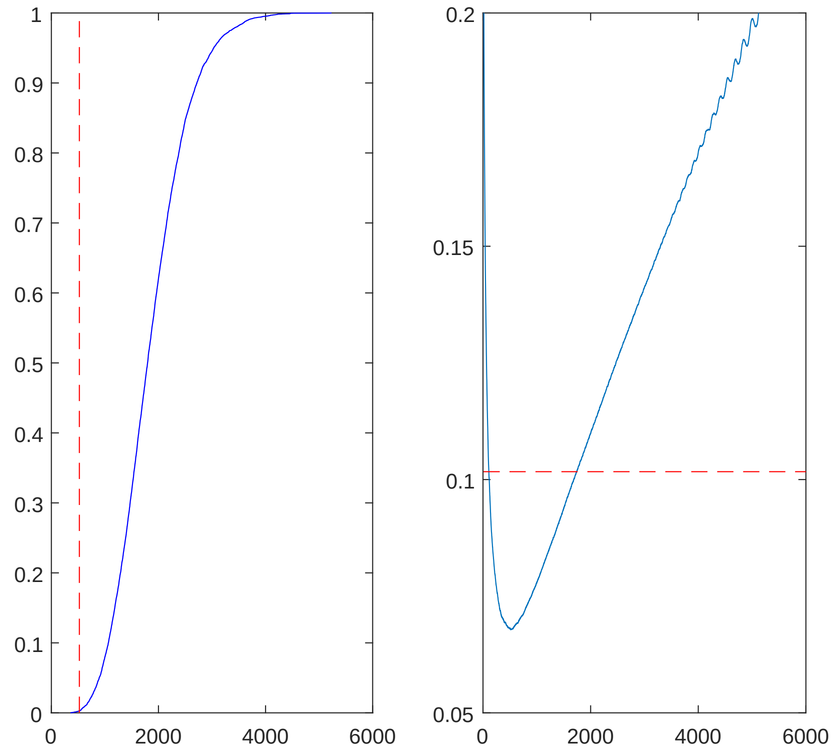


Figure 4.3.2: Left: empirical cdf of k_n^* for the linear PA Model II, the RMSE minimizing value of k is indicated by the dashed red line; right: RMSE of the Hill estimator vs. k, the RMSE of $\hat{\alpha}_{n,k_n^*}$ is indicated by the dashed red line.

RMSE of $\hat{\alpha}_{n,k_n^*}$ is about 50.0% larger than the minimal RMSE. For this model, the RMSE increases much faster as k deviates from the RMSE-minimizing value $k = 523$ (see the right plot of Figure 4.3.2). Since the distribution of k_n^* (with an estimated mean of 1857) puts almost all its mass on values of k much larger than the point of minimum, the sensitivity of the Hill estimator to an inappropriate selection of k leads to a rather poor performance of α_{n,k_n^*}.

Though the performance of the MDSP in our simulation is somewhat mixed, it yields good results in terms of the RMSE of the Hill estimator if the Hill estimator is not very sensitive to the choice of the threshold. According to further simulation results (not reported here), such a behavior seems to be more common for network data than for iid Pareto data. Hence, we conclude that the MDSP often works well on the linear PA models under proper choices of parameters.

4.4 Proofs

4.4.1 Proof of Theorem 4.2.1

Let U_i, $i \in \mathbb{N}$, be iid uniform rv's, ξ_i, $i \in \mathbb{N}$, be iid standard exponential rv's, and $S_k := \sum_{i=1}^k \xi_i$. Then $(U_{(n-k+1)})_{1\le k\le n} =^d (S_k/S_{n+1})_{1\le k\le n}$ for all $n \in \mathbb{N}$ ([48], Cor. 1.6.9). Hence, by the quantile transformation, it suffices to prove that the assertion holds for

$$\tilde{D}_k = \max_{1\le j\le k} \left| \Big(\frac{F^{\leftarrow}(1 - S_j/S_{n+1})}{F^{\leftarrow}(1 - S_k/S_{n+1})} \Big)^{-\tilde{\alpha}_k} - \frac{j}{k} \right| \tag{4.4.1}$$

instead of D_k with

$$\tilde{\alpha}_{n,k} := \Big(\frac{1}{k-1} \sum_{i=1}^{k-1} \log \frac{F^{\leftarrow}(1 - S_i/S_{n+1})}{F^{\leftarrow}(1 - S_k/S_{n+1})} \Big)^{-1}$$

and $F^{\leftarrow}$ denoting the quantile function of F. For F according to (4.2.1) this simplifies to

$$\tilde{D}_k = \max_{1\le j\le k} \left| \Big(\frac{S_j}{S_k} \Big)^{\tilde{\alpha}_{n,k}/\alpha} - \frac{j}{k} \right| \tag{4.4.2}$$

with

$$\frac{\alpha}{\tilde{\alpha}_{n,k}} = \frac{1}{k-1} \sum_{i=1}^{k-1} \log \frac{S_k}{S_i} = \frac{1}{k-1} \sum_{i=1}^{k-1} \log \frac{k}{i} - \frac{1}{k-1} \sum_{i=1}^{k-1} \log \frac{kS_i}{iS_k}. \tag{4.4.3}$$

Note that neither $\tilde{D}_k$ nor $\tilde{\alpha}_{n,k}$ depend on n, so that we will drop the index n when using the latter in the remaining part of the proof.

The first sum on the right hand side of (4.4.3) is a Riemann approximation of $\int_0^1 \log(1/t)\,dt = 1$ with an approximation error $O((\log k)/k)$. To analyze the second sum, we use the so-called Hungarian construction (see [36, 37]): for suitable versions of the ξ_i, there exists a Brownian motion W such that

$$\max_{1\le i\le k} |S_i - i - W(i)| = O(\log k) \quad \text{a.s.}$$

Let $LL_x := \log\log(\mathrm{e}^{\mathrm{e}} \vee x)$. Then

$$\begin{aligned}
\frac{kS_i}{iS_k} - 1 &= \frac{W(i) - (i/k)W(k) + O(\log k)}{i + (i/k)W(k) + O((i/k)\log k)} \\
&= \frac{W(i)}{i} - \frac{W(k)}{k} + O\Big(\frac{\log k}{i} + \Big(\frac{LL_i LL_k}{ik}\Big)^{1/2}\Big) \\
&= \frac{W(i)}{i} - \frac{W(k)}{k} + O\Big(\frac{\log k}{i}\Big) \qquad (4.4.4)
\end{aligned}$$

uniformly for all $1 \le i \le k$, where in the second step we have used the law of iterated logarithm.

It follows by the strong law of large numbers and a Taylor expansion of log that

$$\begin{aligned}
\frac{1}{k-1}\sum_{i=1}^{k-1} \log\frac{kS_i}{iS_k} &= \frac{1}{k-1}\sum_{i=\lceil\log k\rceil}^{k-1} \log\Big[1 + \frac{W(i)}{i} - \frac{W(k)}{k} + O\Big(\frac{\log k}{i}\Big)\Big] + O\Big(\frac{\log k}{k}\Big) \\
&= \frac{1}{k-1}\sum_{i=\lceil\log k\rceil}^{k-1} \Big(\frac{W(i)}{i} - \frac{W(k)}{k} + O\Big(\frac{\log k}{i}\Big)\Big) + O\Big(\frac{\log k}{k}\Big) \\
&= \frac{1}{k-1}\sum_{i=\lceil\log k\rceil}^{k-1} \Big(\frac{W(i)}{i} - \frac{W(k)}{k}\Big) + O\Big(\frac{\log^2 k}{k}\Big),
\end{aligned}$$

since $\sum_{i=1}^k i^{-1} = O(\log k)$. Moreover, by the law of iterated logarithm,

$$\frac{1}{k-1}\sum_{i=1}^{\lceil\log k\rceil-1} \Big(\frac{W(i)}{i} - \frac{W(k)}{k}\Big) = O\Big(\frac{\log k}{k}\Big).$$

To sum up, we have shown that

$$\frac{\alpha}{\tilde{\alpha}_k} - 1 = -\frac{1}{k-1}\sum_{i=1}^{k-1} \Big(\frac{W(i)}{i} - \frac{W(k)}{k}\Big) + O\Big(\frac{\log^2 k}{k}\Big),$$

which in turn implies

$$\frac{\tilde{\alpha}_k}{\alpha} = 1 + \frac{1}{k-1}\sum_{i=1}^{k-1}\Big(\frac{W(i)}{i} - \frac{W(k)}{k}\Big) + O\Big(\frac{\log^2 k}{k}\Big). \tag{4.4.5}$$

Let $\tau_k = \tilde{\alpha}_k/\alpha - 1 = O((LL_k/k)^{1/2})$. Then, by (4.4.4), one has, uniformly for all $1 \le j \le k$,

$$\Big(\frac{S_j}{S_k}\Big)^{\tilde{\alpha}_k/\alpha} = \Big(\frac{j}{k}\Big)^{1+\tau_k}\Big[1 + \frac{W(j)}{j} - \frac{W(k)}{k} + O\Big(\frac{\log k}{j}\Big)\Big]^{1+\tau_k}.$$

The first factor on the right hand side equals

$$\frac{j}{k}\Big(1 + \tau_k \log\frac{j}{k} + O\Big(\frac{\log^2(j/k)LL_k}{k}\Big)\Big).$$

A Taylor expansion of log and exp shows that the second factor is equal to

$$\begin{aligned} &\exp\Big((1+\tau_k)\log\Big[1 + \frac{W(j)}{j} - \frac{W(k)}{k} + O\Big(\frac{\log k}{j}\Big)\Big]\Big) \\ = \;& \exp\Big((1+\tau_k)\Big[\frac{W(j)}{j} - \frac{W(k)}{k} + O\Big(\frac{\log k}{j}\Big)\Big]\Big) \\ = \;& 1 + \frac{W(j)}{j} - \frac{W(k)}{k} + O\Big(\frac{\log k}{j}\Big), \end{aligned}$$

because $\tau_k(LL_j/j)^{1/2} = o((\log k)/j)$. Therefore, since $t(\log t)^2$ is bounded on the unit interval and thus $O(\log^2(j/k)LL_k/k) = o((\log k)/j)$,

$$\Big(\frac{S_j}{S_k}\Big)^{\tilde{\alpha}_k/\alpha} = \frac{j}{k}\Big(1 + \frac{W(j)}{j} - \frac{W(k)}{k} + \tau_k\log\frac{j}{k} + O\Big(\frac{\log k}{j}\Big)\Big) \tag{4.4.6}$$

uniformly for all $1 \le j \le k$. Combining (4.4.2), (4.4.5) and (4.4.6), we arrive at

$$\tilde{D}_k = \max_{1\le j\le k}\Big|\frac{j}{k}\Big(\frac{W(j)}{j} - \frac{W(k)}{k}\Big) + \frac{j}{k}\log\frac{j}{k}\Big(\frac{1}{k-1}\sum_{i=1}^{k-1}\frac{W(i)}{i} - \frac{W(k)}{k}\Big)\Big| + O\Big(\frac{\log^2 k}{k}\Big). \tag{4.4.7}$$

In the last step, we replace the maximum over the discrete points j with a supremum over a whole interval and the sum with an integral. To this end, for each n, we define the Brownian motion $W_n(x) = n^{-1/2}W(nx)$, $x \ge 0$. Then, with

$k = \lceil nt \rceil$ and $j = sk$

$$n^{1/2}\tilde{D}_{\lceil nt\rceil} = \max_{\substack{s\in(0,1]\\ s\lceil nt\rceil\in\mathbb{N}}} \Big| s\Big(\frac{W_n(s\lceil nt\rceil/n)}{s\lceil nt\rceil/n} - \frac{W_n(\lceil nt\rceil/n)}{\lceil nt\rceil/n}\Big) + \\ + s\log s\Big(\frac{1}{\lceil nt\rceil - 1}\sum_{i=1}^{\lceil nt\rceil-1}\frac{W_n(i/n)}{i/n} - \frac{W_n(\lceil nt\rceil/n)}{\lceil nt\rceil/n}\Big)\Big| + O\Big(\frac{\log^2(nt)}{n^{1/2}t}\Big). \quad (4.4.8)$$

Recall that the modulus of continuity of a Brownian motion on the unit interval equals $\omega_W(\delta) = (2\delta|\log\delta|)^{1/2}$ a.s. Hence

$$\sup_{s\in(0,1]}\big|W_n(s\lceil nt\rceil/n) - W_n(st)\big| = O_P\Big(\Big(\frac{\log n}{n}\Big)^{1/2}\Big). \quad (4.4.9)$$

Furthermore, for all $1/n \le x \le y \le 1$

$$\Big|\frac{W_n(y)}{y} - \frac{W_n(x)}{x}\Big| \le \frac{|W_n(y) - W_n(x)|}{y} + |W_n(x)|\frac{|y-x|}{xy}. \quad (4.4.10)$$

Thus, we conclude using the law of iterated logarithm (at 0) that, uniformly for $t \in [2/n, 1]$, $s \in [1/\lceil nt\rceil, 1]$,

$$s\Big|\frac{W_n(s\lceil nt\rceil/n)}{s\lceil nt\rceil/n} - \frac{W_n(st)}{st}\Big| = O_P\Big(\Big(\frac{\log n}{n}\Big)^{1/2}\frac{1}{t} + (stLL_{1/(st)})^{1/2}\frac{s^2/n}{(st)^2}\Big) \\ = O_P\Big(\Big(\frac{\log n}{n}\Big)^{1/2}\frac{1}{t}\Big). \quad (4.4.11)$$

Finally, again using (4.4.10), we obtain

$$\frac{1}{\lceil nt\rceil - 1}\sum_{i=1}^{\lceil nt\rceil-1}\frac{W_n(i/n)}{i/n} - \int_0^1 \frac{W_n(tx)}{tx}\,dx \\ = \sum_{i=1}^{\lceil nt\rceil-1}\int_{(i-1)/\lceil nt-1\rceil}^{i/\lceil nt-1\rceil}\frac{W_n(i/n)}{i/n} - \frac{W_n(tx)}{tx}\,dx \\ = O_P\Big(\frac{1}{\lceil nt\rceil - 1}\Big[\sum_{i=1}^{\lceil nt\rceil-1}\frac{n}{i}\Big(\Big|\log\frac{t}{\lceil nt\rceil-1}\Big|\frac{t}{\lceil nt\rceil-1}\Big)^{1/2} + \\ + \Big(LL_{n/i}\frac{i}{n}\Big)^{1/2}\frac{n}{i}\frac{t}{\lceil nt\rceil - 1}\Big]\Big) \\ = O_P\Big(\Big(\frac{\log n}{n}\Big)^{1/2}\frac{\log(nt)}{t}\Big) \quad (4.4.12)$$

uniformly for $t \in [2/n, 1]$. Combining (4.4.8), (4.4.9), (4.4.11) and (4.4.12), we conclude

$$n^{1/2}\tilde{D}_{\lceil nt\rceil} = \max_{\substack{s\in(0,1]\\ s\lceil nt\rceil\in\mathbb{N}}} \Big| s\Big(\frac{W_n(st)}{st} - \frac{W_n(t)}{t}\Big) + s\log s\Big(\int_0^1 \frac{W_n(tx)}{tx}\,dx - \frac{W_n(t)}{t}\Big)\Big| + O_P\Big(\frac{\log(nt)(\log(nt) + (\log n)^{1/2})}{n^{1/2}t}\Big) \tag{4.4.13}$$

uniformly for $t \in [2/n, 1]$.

To replace the maximum with a supremum over all $s \in (0, 1]$, observe that for any $s \in (0, 1]$ there is a point $\tilde{s}$ with $|s - \tilde{s}| < 1/(nt)$ that is considered in maximum. Hence by the modulus of continuity of W_n, the law of iterated logarithm and the inequality $|s\log s - \tilde{s}\log\tilde{s}| \le |s - \tilde{s}|(1 + |\log(s \wedge \tilde{s})|)$, which holds for all $s, \tilde{s} \in (0, 1]$, one has

$$\begin{aligned}
&\Big| \max_{\substack{s\in(0,1]\\ s\lceil nt\rceil\in\mathbb{N}}} \Big| s\Big(\frac{W_n(st)}{st} - \frac{W_n(t)}{t}\Big) + s\log s\Big(\int_0^1 \frac{W_n(tx)}{tx}\,dx - \frac{W_n(t)}{t}\Big)\Big| \\
&- \sup_{s\in(0,1]} \Big| s\Big(\frac{W_n(st)}{st} - \frac{W_n(t)}{t}\Big) + s\log s\Big(\int_0^1 \frac{W_n(tx)}{tx}\,dx - \frac{W_n(t)}{t}\Big)\Big|\Big| \\
&= O_P\Big(\frac{1}{t}\Big(\frac{\log n}{n}\Big)^{1/2} + \frac{1}{nt}\Big(\frac{LL_{1/t}}{t}\Big)^{1/2}\Big) + O_P\Big(\frac{1}{nt}(1 + \log(nt))\Big(\frac{LL_{1/t}}{t}\Big)^{1/2}\Big).
\end{aligned}$$

Now the assertion follows readily.

4.4.2 Proof of Corollary 4.2.2

By Theorem 4.2.1, $k^{1/2}D_k \to \sup_{s\in(0,1]}|Z_1(s,1)|$ weakly as $k \to \infty$. According to Skohorod's theorem, there exist versions such that the convergence holds almost surely. Moreover, all D_k and the limit random variable as well are almost surely strictly positive. Hence, for all sequences $k_n = o(n)$, it follows that $n^{1/2}\min_{2\le k\le k_n} D_k \ge (n/k_n)^{1/2}\min_{2\le k\le k_n} k^{1/2}D_k \to \infty$ almost surely. This implies that with probability tending to 1, k_n^* must be larger than k_n.

Because $k_n = o(n)$ is arbitrary, we conclude that the sequence n/k_n^* is stochastically bounded. Since, by Theorem 4.2.1, $(n^{1/2}D_{\lceil nt \rceil})_{t\in[\varepsilon,1]}$ converges weakly to $(Z_1(t))_{t\in[\varepsilon,1]}$ (w.r.t. the supremum norm) for all $\varepsilon > 0$ and Z_1 is continuous on $(0, 1]$ with a unique point of minimum, the asymptotic behavior of k_n^*/n follows (cf. Corollary 5.58 of [60]). The asymptotics of the Hill estimator can be easily derived from (4.4.5) and the approximations established in the last part of the proof of Theorem 4.2.1, in particular (4.4.9) and (4.4.12).

4.5 Conclusions

We discussed the asymptotic and the finite sample performance of the minimum distance selection procedure. It was shown that, unlike previously proposed methods, the sample fraction k_n^*/n chosen by the MDSP does not asymptotically concentrate on one point. Instead, it often yields too small a value of k, leading to a strongly increased variance and RMSE of the Hill estimator.

As the simulations of linear preferential attachment networks have shown, the spread of the distribution of k_n^* need not always result in a large loss of efficiency of the Hill estimator. This is particularly true if the Hill estimator is rather insensitive to the choice of k over a wide range of k, because the increase in the bias with growing k is balanced by the decrease of the variance.

There may be another reason why the MDSP shows a considerably different behavior for the in-degrees of linear preferential attachment networks and for iid observations. In the latter situation, for any fixed sample size, all observations are drawn from a distribution which is regularly varying, that is $(1 - F(tx))/(1 - F(x)) \to t^{-1/\alpha}$ as $x \to \infty$. This is not true for the in-degrees of a linear PA model, which cannot take on values larger than the number of edges in the network. In fact, the distribution of the in-degrees changes when the sample size (i.e., the number of nodes) increases, and only in the limit the distribution has a power tail behavior. So while in the situation considered in the Chapter 4.2 the tail index α has the same operational meaning for each sample size, for linear PA models it is defined only as a limit parameter as the number n of nodes tends to infinity. Since, strictly speaking, for any fixed n, there is no tail index, the interpretation of the RMSE of the Hill estimator is somewhat unclear in this setting.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

In this thesis, we have been focusing on modeling the power-law behavior in directed linear preferential attachment networks, where both statistical and probabilistic analyses have been conducted.

From the statistical point of view, we consider methods that can fit the preferential attachment model under different data scenarios. If the full history of the network evolution is available, we obtain maximum likelihood estimates (MLE's) of the model parameters and these MLE's are strongly consistent, asymptotically normal and asymptotically efficient. If only a single snapshot of the network is given, we derive snapshot estimates by combining the method of moments with an approximation to the likelihood. In Chapter 2.3, we show that snapshot estimates are also strongly consistent and their asymptotic normality is checked through simulation studies. The loss of efficiency for the snapshot estimates turns out to be less than one would expect given the substantial reduction in the data available to produce the snapshot estimates.

When fitting the preferential attachment model to real data, we realize that in practice, datasets may be corrupted. Simulation results affirm that likelihood-based parametric estimation approaches lead to biased estimates so are not robust against data corruption. To accommodate this, we propose another semi-parametric approach which uses extreme value techniques. We compare results from the extreme value estimation method and those from MLE and snapshot methods across different levels of data corruption and see that the extreme value method is more robust than the other two.

Even though the power-law tail index of the degree distribution is often estimated through the Hill estimator, no prior rigorous justification on its consistency has been provided in the network content. We study the joint growth of the in- and out-degrees in a linear preferential attachment model and prove the consistency of the Hill estimator for network data.

Another issue with the Hill estimation is how to determine the threshold above which the distribution is close to a power law. A widely adopted procedure[10] had no justification. We analyze this threshold selection procedure in the simple iid Pareto case and find that asymptotic distributions of the threshold selected and the corresponding tail index estimate are very complicated. This makes the statistical inference harder. We also see that this threshold selection method gives estimates with large MSE in the Pareto case but performs better in the network content.

There are several open research questions related to the work in this thesis. The first is to examine the asymptotic distribution of the Hill estimator in the preferential attachment model. We have shown the asymptotic normality of degree counts in [66] and whether it can be used to prove the asymptotic normality of the Hill estimator remains unknown. With the asymptotic distribution available, more work on formal statistical inferences can be done, e.g. constructing confidence intervals for the estimated tail exponent and testing for any deviations from the preferential attachment model. Simulation evidence suggests the asymptotic normality and the asymptotic variance of the Hill estimate turns out to be smaller than one would have in the iid case. This may be due to the complicated dependence structure in the network setup and may partly explain why the threshold selection method in Chapter 4 gives estimates with lower

MSE.

Meanwhile, the extreme value estimation method outlined in Chapter 2.6 has not been applied to real datasets yet. In practice, it is hard to justify whether the real dataset has been corrupted since we do not know what the true model is. One possible research project is to try fitting the preferential attachment model to some real dataset which is believed to have followed a different model, for example, the superstar preferential attachment model studied in [6]. It would be interesting to get some fresh data using the Twitter API and Python libraries like Twython and try fitting the preferential attachment model using the extreme value estimation approach.

There are plenty of other interesting research topics in modeling the social network content, where we can apply some analytical tools discussed in this thesis. For instance, we can apply the embedding technique in Chapter 3.3 to deal with problems like the number of common nodes and reciprocity in a social network like LinkedIn. Also, given the timestamp information of the edge creation in a network, can we model the network growth by some self-exciting process and make prediction on the growth rate of the network? We leave these as future research directions.

APPENDIX A

APPENDICES FOR CHAPTER 2

A.1 Simulation algorithm

We describe an efficient simulation procedure for the preferential attachment network given the parameter values $(\alpha, \beta, \gamma, \delta_{\text{in}}, \delta_{\text{out}})$, where $\alpha + \beta + \gamma = 1$. The simulation cost of the algorithm is linear in time. This algorithm, which was provided by Joyjit Roy during his graduate work at Cornell University, is presented below for completeness. Note that this simulation algorithm is specifically designed for the case where the preferential attachment probabilities (1.1.1)–(1.1.2) are linear in the degrees. A similar idea for the simulation of the Yule-Simon process appeared in [58]. Efficient simulation methods for the case where the preferential attachment probabilities are non-linear are studied in [3], where their algorithm trades some efficiency for the flexibility to model non-linear preferential attachment.

Using the notation from the introduction, at time $t = 0$, we initiate with an arbitrary graph $G(n_0) = (V(n_0), E(n_0))$ of n_0 edges, where the elements of $E(n_0)$ are represented in form of $(v_i^{(1)}, v_i^{(2)}) \in V(n_0) \times V(n_0)$, $i = 1, \ldots, n_0$, with $v_i^{(1)}, v_i^{(2)}$ denoting the outgoing and incoming vertices of the edge, respectively. To grow the network, we update the network at each stage from $G(n-1)$ to $G(n)$ by adding a new edge $(v_n^{(1)}, v_n^{(2)})$. Assume that the nodes are labeled using positive integers starting from 1 according to the time order in which they are created, and let the random number $N(n) = |V(n)|$ denote the total number of nodes in $G(n)$.

Let us consider the situation where an existing node is to be chosen from

Algorithm 1: Simulating a directed edge preferential attachment network

Algorithm

Input: $\alpha, \beta, \delta_{\text{in}}, \delta_{\text{out}}$, the parameter values; $G(n_0) = (V(n_0), E(n_0))$, the initialization graph; n, the targeted number edges

Output: $G(n) = (V(n), E(n))$, the resulted graph

```
t ← n_0
while t < n do
    N(t) ← |V(t)|
    Generate U ~ Uniform(0, 1)
    if U < α then
        v^(1) ← N(t) + 1
        v^(2) ← Node_Sample(E(t), 2, δ_in)
        V(t) ← Append(V(t), N(t) + 1)
    else if α ≤ U < α + β then
        v^(1) ← Node_Sample(E(t), 1, δ_out)
        v^(2) ← Node_Sample(E(t), 2, δ_in)
    else if U ≥ α + β then
        v^(1) ← Node_Sample(E(t), 1, δ_out)
        v^(2) ← N(t) + 1
        V(t) ← Append(V(t), N(t) + 1)
    E(t + 1) ← Append(E(t), (v^(1), v^(2)))
    t ← t + 1
end
return G(n) = (V(n), E(n))
```

Function *Node_Sample*

Input: $E(t)$, the edge list up to time t; $j = 1, 2$, the node to be sample, representing outgoing and incoming nodes, respectively; $\delta \in \{\delta_{\text{in}}, \delta_{\text{out}}\}$, the offset parameter

Output: the sampled node, v

```
Generate W ~ Uniform(0, t + N(t)δ)
if W ≤ t then
    v ← v^(j)_⌈W⌉
else if W > t then
    v ← ⌈(W − t)/δ⌉
return v
```

$V(n)$ as the vertex of the new edge. Naively sampling from the multinomial distribution requires $O(N(n))$ evaluations, where $N(n)$ increases linearly with n. Therefore the total cost to simulate a network of n edges is $O(n^2)$. This is significantly burdensome when n is large, which is usually the case for observed networks. Algorithm 1 describes a simulation algorithm which uses the alias method [41] for node sampling. Here sampling an existing node from $V(n)$ requires only constant execution time, regardless of n. Hence the cost to simulate $G(n)$ is only $O(n)$. This method allows generation of a graph with 10^7 nodes on a personal laptop in less than 5 seconds.

To see that the algorithm indeed produces the intended network, it suffices to consider the case of sampling an existing node from $V(n-1)$ as the incoming vertex of the new edge. In the function Node_Sample in Algorithm 1, we generate $W \sim \text{Uniform}(0, n-1+N(n-1)\delta_{\text{in}})$ and set

$$v \leftarrow v^{(j)}_{\lceil W \rceil} \mathbf{1}_{\{W \leq n-1\}} + \left\lceil \frac{W-(n-1)}{\delta_{\text{in}}} \right\rceil \mathbf{1}_{\{W > n-1\}}.$$

Then

$$\begin{aligned}
\mathbf{P}(v = w) &= \mathbf{P}\left(v^{(j)}_{\lceil W \rceil} = w\right) \mathbf{P}(W \leq n-1) \\
&\quad + \mathbf{P}\left(\left\lceil \frac{W-(n-1)}{\delta_{\text{in}}} \right\rceil = w\right) \mathbf{P}(W > n-1) \\
&= \frac{D^{(n-1)}_{\text{in}}(w)}{n-1} \frac{n-1}{n-1+N(n-1)\delta_{\text{in}}} \\
&\quad + \frac{1}{N(n-1)} \frac{N(n-1)\delta_{\text{in}}}{n-1+N(n-1)\delta_{\text{in}}} \\
&= \frac{D^{(n-1)}_{\text{in}}(w) + \delta_{\text{in}}}{n-1+N(n-1)\delta_{\text{in}}},
\end{aligned}$$

which corresponds to the desired selection probability (1.1.1).

A.2 For the proof of Theorem 2.2.2: Lemmas A.2.1 and A.2.2

Lemma A.2.1. *For $\lambda > 0$, the function $\psi(\lambda)$ in* (2.2.10) *has a unique zero at δ_{in} and, $\psi(\lambda) > 0$ when $\lambda < \delta_{in}$ and $\psi(\lambda) < 0$ when $\lambda > \delta_{in}$.*

Proof. The probabilities $\{p_i^{\text{in}}(\lambda)\}$ satisfy the recursions in i (cf. [7]):

$$\begin{aligned}
p_0^{\text{in}}(\lambda)\left(\lambda + \frac{1}{a_1(\lambda)}\right) &= \frac{\alpha}{a_1(\lambda)}, && \text{(A.2.1a)}\\
p_1^{\text{in}}(\lambda)\left(1 + \lambda + \frac{1}{a_1(\lambda)}\right) &= \lambda p_0^{\text{in}}(\lambda) + \frac{\gamma}{a_1(\lambda)}, \\
p_2^{\text{in}}(\lambda)\left(2 + \lambda + \frac{1}{a_1(\lambda)}\right) &= (1+\lambda) p_1^{\text{in}}(\lambda), \\
&\vdots \\
p_i^{\text{in}}(\lambda)\left(i + \lambda + \frac{1}{a_1(\lambda)}\right) &= (i-1+\lambda) p_{i-1}^{\text{in}}(\lambda), \quad (i \geq 2),
\end{aligned}$$

where $a_1(\lambda) := (\alpha+\beta)/(1+\lambda(1-\beta))$. Summing the recursions in (A.2.1) from 0 to i, we get (with the convention that $\sum_{i=0}^{-1} = 0$)

$$\sum_{k=0}^{i} p_k^{\text{in}}(\lambda)\left(k + \lambda + \frac{1}{a_1(\lambda)}\right) = \sum_{k=0}^{i-1}(k+\lambda)p_k^{\text{in}}(\lambda) + \frac{\alpha}{a_1(\lambda)} + \frac{\gamma}{a_1(\lambda)}\mathbf{1}_{\{i\geq 1\}}, \quad i \geq 0,$$

which can be simplified to

$$\frac{1}{a_1(\lambda)}\sum_{k=0}^{i} p_k^{\text{in}}(\lambda) + (i+\lambda)p_i^{\text{in}}(\lambda) = \frac{1-\beta}{a_1(\lambda)} - \frac{\gamma}{a_1(\lambda)}\mathbf{1}_{\{i=0\}}, \quad i \geq 0. \tag{A.2.2}$$

From (1.1.3),

$$\sum_{i=0}^{\infty} p_i^{\text{in}}(\lambda) = \sum_{i,j} p_{ij}(\lambda) = 1-\beta. \tag{A.2.3}$$

Hence by rearranging (A.2.2), we have

$$(i+\lambda)p_i^{\text{in}}(\lambda) + \frac{\gamma}{a_1(\lambda)}\mathbf{1}_{\{i=0\}} = \frac{1}{a_1(\lambda)}\left(1-\beta-\sum_{k=0}^{i} p_k^{\text{in}}(\lambda)\right) = \frac{1}{a_1(\lambda)}p_{>i}^{\text{in}}(\lambda),$$

or equivalently,

$$p_{>i}^{\text{in}}(\lambda) = a_1(\lambda)(i+\lambda)p_i^{\text{in}}(\lambda) + \gamma\mathbf{1}_{\{i=0\}}. \tag{A.2.4}$$

Now with the help of (A.2.3) and (A.2.4), we can rewrite $\psi(\lambda)$ in the following way:

$$
\begin{aligned}
\psi(\lambda) &= \sum_{i=0}^{\infty} \frac{p_{>i}^{\text{in}}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - (1-\beta)a_1(\lambda) \\
&= \sum_{i=0}^{\infty} \frac{p_{>i}^{\text{in}}(\delta_{\text{in}})}{i+\lambda} - \frac{\gamma}{\lambda} - \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})a_1(\lambda)(i+\lambda)}{i+\lambda} \\
&= \sum_{i=0}^{\infty} \frac{a_1(\delta_{\text{in}})(i+\delta_{\text{in}})p_i^{\text{in}}(\delta_{\text{in}}) + \gamma \mathbf{1}_{\{i=0\}}}{i+\lambda} - \frac{\gamma}{\lambda} - \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})a_1(\lambda)(i+\lambda)}{i+\lambda} \\
&= \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}\Big(a_1(\delta_{\text{in}})(i+\delta_{\text{in}}) - a_1(\lambda)(i+\lambda)\Big) \\
&= \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda} \int_{\lambda}^{\delta_{\text{in}}} \frac{\partial}{\partial s}\Big(a_1(s)(i+s)\Big)ds \\
&= \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda} \int_{\lambda}^{\delta_{\text{in}}} \frac{(\alpha+\beta)(1-i(1-\beta))}{(1+s(1-\beta))^2}ds \\
&= \left(\sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}(1-i(1-\beta))\right) \int_{\lambda}^{\delta_{\text{in}}} \frac{\alpha+\beta}{(1+s(1-\beta))^2}ds \\
&=: C(\lambda) \int_{\lambda}^{\delta_{\text{in}}} \frac{\alpha+\beta}{(1+s(1-\beta))^2}ds. \qquad \text{(A.2.5)}
\end{aligned}
$$

The series defining $C(\lambda)$ converges absolutely for any $\lambda > 0$ since

$$
\begin{aligned}
\sum_{i=0}^{\infty} \left| \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}(1-i(1-\beta)) \right| &< \sum_{i=0}^{\infty} p_i^{\text{in}}(\delta_{\text{in}}) \left| \frac{i(1-\beta)}{i+\lambda} + \frac{1}{i+\lambda} \right| \\
&< (1-\beta)(1-\beta+\frac{1}{\lambda}) < \infty.
\end{aligned}
$$

Summing over i in (A.2.4), we get by monotone convergence

$$
\sum_{i=0}^{\infty} p_{>i}^{\text{in}}(\lambda) = \sum_{i=0}^{\infty} i p_i^{\text{in}}(\lambda) = a_1(\lambda) \sum_{i=0}^{\infty} i p_i^{\text{in}}(\lambda) + a_1(\lambda)\lambda \sum_{i=0}^{\infty} p_i^{\text{in}}(\lambda) + \gamma.
$$

The infinite series converge because $p_i^{\text{in}}(\lambda)$ is a power law with index greater than 2; see (1.1.4) and (1.1.5). Solving for the infinite series we get

$$
\sum_{i=0}^{\infty} i p_i^{\text{in}}(\lambda) = \frac{a_1(\lambda)\lambda}{1-a_1(\lambda)}(1-\beta) + \frac{\gamma}{1-a_1(\lambda)} = 1. \qquad \text{(A.2.6)}
$$

Hence we have

$$
\begin{aligned}
C(\lambda) &= \sum_{i \le (1-\beta)^{-1}} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}(1-i(1-\beta)) - \sum_{i>(1-\beta)^{-1}} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{i+\lambda}(i(1-\beta)-1) \\
&> \sum_{i=0}^{\infty} \frac{p_i^{\text{in}}(\delta_{\text{in}})}{(1-\beta)^{-1}+\lambda}(1-i(1-\beta)) \\
&= \frac{1}{(1-\beta)^{-1}+\lambda} \sum_{i=0}^{\infty} p_i^{\text{in}}(\delta_{\text{in}}) - \frac{1-\beta}{(1-\beta)^{-1}+\lambda} \sum_{i=0}^{\infty} i p_i^{\text{in}}(\delta_{\text{in}}) \\
&= \frac{1}{(1-\beta)^{-1}+\lambda}(1-\beta) - \frac{1-\beta}{(1-\beta)^{-1}+\lambda} 1 \\
&= 0.
\end{aligned}
$$

Now recall from (A.2.5) that $\psi(\lambda)$ is of the form

$$
\psi(\lambda) = C(\lambda) \int_{\lambda}^{\delta_{\text{in}}} \frac{\alpha+\beta}{(1+s(1-\beta))^2} ds,
$$

where $C(\lambda) > 0$ for all $\lambda > 0$. Therefore $\psi(\cdot)$ has a unique zero at δ_{in} and $\psi(\lambda) > 0$ when $\lambda < \delta_{\text{in}}$ and $\psi(\lambda) < 0$ when $\lambda > \delta_{\text{in}}$. □

We show the uniform convergence of ψ_n to ψ in the next lemma.

Lemma A.2.2. *As $n \to \infty$, for any $\epsilon > 0$,*

$$
\sup_{\lambda \ge \epsilon} |\psi_n(\lambda) - \psi(\lambda)| \xrightarrow{a.s.} 0.
$$

Proof. By the definition of ψ, $p_{>i}^{\text{in}}(\delta_{\text{in}})$ is a function of δ_{in} and is a constant with respect to λ. Hence we suppress the dependence on δ_{in} and simply write it as $p_{>i}^{\text{in}}$ when considering the difference $\psi_n - \psi$ as a function of λ:

$$
\begin{aligned}
\psi_n(\lambda) - \psi(\lambda) &= \sum_{i=0}^{\infty} \frac{N_{>i}^{\text{in}}(n)/n - p_{>i}^{\text{in}}}{i+\lambda} - \frac{1}{\lambda}\left(\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}} - (1-\alpha-\beta)\right) \\
&\quad - \frac{1}{n}\sum_{t=1}^{n}\left(\frac{N(t-1)}{t-1+\lambda N(t-1)} \mathbf{1}_{\{J_t \in \{1,2\}\}} - \frac{(1-\beta)(\alpha+\beta)}{1+\lambda(1-\beta)}\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\sup_{\lambda\ge\epsilon}|\psi_n(\lambda)-\psi(\lambda)| \\
\le\ &\sup_{\lambda\ge\epsilon}\sum_{i=0}^{\infty}\frac{\left|N^{\text{in}}_{>i}(n)/n-p^{\text{in}}_{>i}\right|}{i+\lambda}+\sup_{\lambda\ge\epsilon}\frac{1}{\lambda}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=3\}}-(1-\alpha-\beta)\right| \\
&+\sup_{\lambda\ge\epsilon}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{N(t-1)}{t-1+\lambda N(t-1)}\mathbf{1}_{\{J_t\in\{1,2\}\}}-\frac{(1-\beta)(\alpha+\beta)}{1+\lambda(1-\beta)}\right|.
\end{aligned}
\tag{A.2.7}
$$

For the first term, note that for all $i\ge 0$,

$$
iN^{\text{in}}_{>i}(n)=\sum_{k=i+1}^{\infty}N^{\text{in}}_{k}(n)i\le\sum_{k=1}^{\infty}kN^{\text{in}}_{k}(n)=n,
$$

since the assumption on initial conditions implies the sum of in-degrees at n is n. Therefore $N^{\text{in}}_{>i}(n)/n\le i^{-1}$ for $i\ge 1$, and it then follows that

$$
\sum_{i=0}^{\infty}\frac{\left|N^{\text{in}}_{>i}(n)/n-p^{\text{in}}_{>i}\right|}{i+\lambda}\le\sum_{i=0}^{M}\frac{\left|N^{\text{in}}_{>i}(n)/n-p^{\text{in}}_{>i}\right|}{i+\lambda}+\sum_{i=M+1}^{\infty}\frac{1/i}{i+\lambda}+\sum_{i=M+1}^{\infty}\frac{p^{\text{in}}_{>i}}{i+\lambda}.
$$

Note that the last two terms on the right side can be made arbitrarily small uniformly on $[\epsilon,\infty)$ if we choose M sufficiently large.

Recall the convergence of the degree distribution $\{N_{ij}(n)/N(n)\}$ to the probability distribution $\{f_{ij}\}$ in (1.1.3), we have

$$
\frac{N^{\text{in}}_{>i}(n)}{n}=\frac{N(n)}{n}\frac{N^{\text{in}}_{>i}(n)}{N(n)}\xrightarrow{\text{a.s.}}(1-\beta)\sum_{l\ge0,k>i}f_{kl}=p^{\text{in}}_{>i},\quad\forall i\ge0.
\tag{A.2.8}
$$

Hence, for any fixed M,

$$
\sum_{i=0}^{M}\frac{\left|N^{\text{in}}_{>i}(n)/n-p^{\text{in}}_{>i}\right|}{i+\epsilon}\xrightarrow{\text{a.s.}}0,\quad\text{as } n\to\infty.
$$

which implies further that choosing M arbitrarily large gives

$$
\begin{aligned}
&\sup_{\lambda\ge\epsilon}\sum_{i=0}^{\infty}\frac{\left|N^{\text{in}}_{>i}(n)/n-p^{\text{in}}_{>i}\right|}{i+\lambda} \\
&\qquad\le\sum_{i=0}^{M}\frac{\left|N^{\text{in}}_{>i}(n)/n-p^{\text{in}}_{>i}\right|}{i+\epsilon}+\sum_{i=M+1}^{\infty}\frac{1/i}{i+\epsilon}+\sum_{i=M+1}^{\infty}\frac{p^{\text{in}}_{>i}}{i+\epsilon}\xrightarrow{\text{a.s.}}0.
\end{aligned}
$$

The second term in (A.2.7) converges to 0 almost surely by strong law of large numbers, and the third term in (A.2.7) can be written as

$$\left|\frac{1}{n}\sum_{t=1}^{n}\left(\frac{N(t-1)}{t-1+\lambda N(t-1)}-\frac{(1-\beta)}{1+\lambda(1-\beta)}\right)\mathbf{1}_{\{J_t\in\{1,2\}\}}\right.$$
$$\left.+\frac{1-\beta}{1+\lambda(1-\beta)}\frac{1}{n}\sum_{t=1}^{n}(\mathbf{1}_{\{J_t\in\{1,2\}\}}-(\alpha+\beta))\right|,$$

which is bounded by

$$\left|\frac{1}{n}\sum_{t=1}^{n}\frac{N(t-1)}{t-1+\lambda N(t-1)}-\frac{(1-\beta)}{1+\lambda(1-\beta)}\right|$$
$$+\frac{1-\beta}{1+\lambda(1-\beta)}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}}-(\alpha+\beta)\right|.$$

We have

$$\begin{aligned}
&\sup_{\lambda\geq\epsilon}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{N(t-1)}{t-1+\lambda N(t-1)}-\frac{(1-\beta)}{1+\lambda(1-\beta)}\right| \\
&\quad= \sup_{\lambda\geq\epsilon}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{N(t-1)/(t-1)-(1-\beta)}{(1+\lambda N(t-1)/(t-1))(1+\lambda(1-\beta))}\right| \\
&\quad\leq \frac{1}{n}\sum_{t=1}^{n}\left|\frac{N(t-1)/(t-1)-(1-\beta)}{(1+\epsilon N(t-1)/(t-1))(1+\epsilon(1-\beta))}\right|,
\end{aligned}$$

which converges to 0 almost surely by Cesàro convergence of random variables, since

$$\left|\frac{N(n)/n-(1-\beta)}{(1+\epsilon N(n)/n)(1+\epsilon(1-\beta))}\right| \stackrel{\text{a.s.}}{\longrightarrow} 0,\ \text{ as } n\to\infty.$$

Further, by the strong law of large numbers,

$$\begin{aligned}
&\sup_{\lambda\geq\epsilon}\frac{1-\beta}{1+\lambda(1-\beta)}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}}-(\alpha+\beta)\right| \\
&\quad\leq \frac{1-\beta}{1+\epsilon(1-\beta)}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}}-(\alpha+\beta)\right| \stackrel{\text{a.s.}}{\longrightarrow} 0,\ \text{ as } n\to\infty.
\end{aligned}$$

Hence the third term of (A.2.7) also goes to 0 almost surely as $n\to\infty$. The result of the lemma follows. □

A.3 For the proof of Theorem 2.2.3: Lemmas A.3.1 and A.3.2

Lemma A.3.1. *As $n \to \infty$,*

$$n^{-1/2} \sum_{t=1}^{n} u_t(\delta_{in}) \xrightarrow{d} N(0, I_{in}). \tag{A.3.1}$$

Proof. Let $\mathcal{F}_n = \sigma(G(0), \ldots, G(n))$ be the σ-field generated by the information contained in the graphs. We first observe that $\{\sum_{t=1}^{n} u_t(\delta_{\text{in}}), \mathcal{F}_n, n \geq 1\}$ is a martingale. To see this, note from (2.2.15) that $|u_t(\delta)| \leq 2/\delta$ and

$$\begin{aligned}
&\mathbf{E}[u_t(\delta_{\text{in}})|\mathcal{F}_{t-1}] \\
&= \mathbf{E}\left[\frac{1}{D^{\text{in}}_{(t-1)}(w_t) + \delta_{\text{in}}} \mathbf{1}_{\{J_t \in \{1,2\}\}} \middle| \mathcal{F}_{t-1} \right] \\
&\quad - \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \mathbf{E}[\mathbf{1}_{\{J_t \in \{1,2\}\}}|\mathcal{F}_{t-1}] \\
&= \mathbf{E}\left[\frac{1}{D^{\text{in}}_{(t-1)}(w_t) + \delta_{\text{in}}} \middle| J_t = 1, \mathcal{F}_{t-1} \right] \mathbf{P}[J_t = 1] \\
&\quad + \mathbf{E}\left[\frac{1}{D^{\text{in}}_{(t-1)}(w_t) + \delta_{\text{in}}} \middle| J_t = 2, \mathcal{F}_{t-1} \right] \mathbf{P}[J_t = 2] \\
&\quad - (\alpha + \beta) \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \\
&= (\alpha + \beta) \sum_{v \in V_{t-1}} \frac{1}{D^{\text{in}}_{(t-1)}(v) + \delta_{\text{in}}} \frac{D^{\text{in}}_{(t-1)}(v) + \delta_{\text{in}}}{t-1+\delta_{\text{in}} N(t-1)} \\
&\quad - (\alpha + \beta) \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \\
&= (\alpha + \beta) \left(\sum_{v \in V_{t-1}} \frac{1}{t-1+\delta_{\text{in}} N(t-1)} - \frac{N(t-1)}{t-1+\delta_{\text{in}} N(t-1)} \right) \\
&= 0,
\end{aligned}$$

which satisfies the definition of a martingale difference. Hence

$$\left\{ n^{-1/2} \sum_{r=1}^{t} u_r(\delta_{\text{in}}) \right\}_{t=1,\ldots,n}$$

is a zero-mean, square-integrable martingale array. The convergence (A.3.1) follows from the martingale central limit theory (cf. Theorem 3.2 of [27]) if the following three conditions can be verified:

(a) $n^{-1/2}\max_t |u_t(\delta_{\mathrm{in}})| \xrightarrow{p} 0$,

(b) $n^{-1}\sum_t u_t^2(\delta_{\mathrm{in}}) \xrightarrow{p} I_{\mathrm{in}}$,

(c) $\mathbf{E}\left(n^{-1}\max_t u_t^2(\delta_{\mathrm{in}})\right)$ is bounded in n.

Since $|u_t(\delta_{\mathrm{in}})| \leq 2/\delta_{\mathrm{in}}$, we have

$$n^{-1/2}\max_t |u_t(\delta_{\mathrm{in}})| \leq \frac{2}{n^{1/2}\delta_{\mathrm{in}}} \to 0,$$

and

$$n^{-1}\max_t u_t^2 \leq \frac{4}{n\delta_{\mathrm{in}}^2} \to 0.$$

Hence conditions (a) and (c) are straightforward.

To show (b), observe that

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} u_t^2(\delta_{\mathrm{in}})\\
&\quad= \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t\in\{1,2\}\}}\left(\frac{1}{D^{\mathrm{in}}_{(t-1)}(w_t)+\delta_{\mathrm{in}}} - \frac{N(t-1)}{t-1+\delta_{\mathrm{in}}N(t-1)}\right)^2\\
&\quad= \frac{1}{n}\sum_{t=1}^{n} \frac{\mathbf{1}_{\{J_t\in\{1,2\}\}}}{\left(D^{\mathrm{in}}_{(t-1)}(w_t)+\delta_{\mathrm{in}}\right)^2}\\
&\qquad - \frac{2}{n}\sum_{t=1}^{n} \frac{\mathbf{1}_{\{J_t\in\{1,2\}\}}}{D^{\mathrm{in}}_{(t-1)}(w_t)+\delta_{\mathrm{in}}}\frac{N(t-1)}{t-1+\delta_{\mathrm{in}}N(t-1)}\\
&\qquad\quad + \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t\in\{1,2\}\}}\left(\frac{N(t-1)}{t-1+\delta_{\mathrm{in}}N(t-1)}\right)^2\\
&\quad =: T_1 - 2T_2 + T_3.
\end{aligned}$$

Following the calculations in the proof of Lemma A.2.2, we have for T_1,

$$T_1 = \sum_{i=0}^{\infty} \frac{N^{\text{in}}_{>i}(n)/n}{(i+\delta_{\text{in}})^2} - \frac{1}{\delta_{\text{in}}^2}\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t=3\}} \xrightarrow{p} \sum_{i=0}^{\infty} \frac{p^{\text{in}}_{>i}}{(i+\delta_{\text{in}})^2} - \frac{\gamma}{\delta_{\text{in}}^2}.$$

We then rewrite T_2 as

$$\begin{aligned}
T_2 &= \frac{1}{n}\sum_{t=1}^{n} \frac{\mathbf{1}_{\{J_t\in\{1,2\}\}}}{D^{\text{in}}_{(t-1)}(w_t)+\delta_{\text{in}}}\left(\frac{N(t-1)/(t-1)}{1+\delta_{\text{in}}N(t-1)/(t-1)} - \frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)}\right)\\
&\quad + \frac{1}{n}\sum_{t=1}^{n} \frac{\mathbf{1}_{\{J_t\in\{1,2\}\}}}{D^{\text{in}}_{(t-1)}(w_t)+\delta_{\text{in}}}\frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)}\\
&=: T_{21}+T_{22},
\end{aligned}$$

where

$$|T_{21}| \le \frac{1}{n}\sum_{t=1}^{n}\frac{1}{\delta_{\text{in}}}\left|\frac{N(t-1)/(t-1)}{1+\delta_{\text{in}}N(t-1)/(t-1)} - \frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)}\right| \xrightarrow{p} 0$$

by Cesàro's convergence and

$$\begin{aligned}
T_{22} &= \frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)}\left(\sum_{i=0}^{\infty}\frac{N^{\text{in}}_{>i}(n)/n}{i+\delta_{\text{in}}} - \frac{1}{\delta_{\text{in}}}\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t=3\}}\right)\\
&\xrightarrow{p} \frac{1-\beta}{1+\delta_{\text{in}}(1-\beta)}\left(\sum_{i=0}^{\infty}\frac{p^{\text{in}}_{>i}}{i+\delta_{\text{in}}} - \frac{\gamma}{\delta_{\text{in}}}\right) = \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\text{in}}(1-\beta))^2},
\end{aligned}$$

where the equality follows from (A.2.4). For T_3, similar to T_1, we have

$$\begin{aligned}
T_3 &= \frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}}\left(\left(\frac{N(t-1)/(t-1)}{1+\delta_{\text{in}}N(t-1)/(t-1)}\right)^2 - \frac{(1-\beta)^2}{(1+\delta_{\text{in}}(1-\beta))^2}\right)\\
&\quad + \frac{(1-\beta)^2}{(1+\delta_{\text{in}}(1-\beta))^2}\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\{J_t\in\{1,2\}\}} \xrightarrow{p} \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\text{in}}(1-\beta))^2}.
\end{aligned}$$

Combining these results together,

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}u_t^2(\delta_{\text{in}}) &= T_1 - 2(T_{21}+T_{22}) + T_3\\
&\xrightarrow{p} \sum_{i=0}^{\infty}\frac{p^{\text{in}}_{>i}}{(i+\delta_{\text{in}})^2} - \frac{\gamma}{\delta_{\text{in}}^2} - \frac{(\alpha+\beta)(1-\beta)^2}{(1+\delta_{\text{in}}(1-\beta))^2} = I_{\text{in}}. \qquad \text{(A.3.2)}
\end{aligned}$$

This completes the proof. □

Lemma A.3.2. *As $n \to \infty$,*

$$\frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\hat{\delta}^*_{in}) \xrightarrow{p} -I_{in}.$$

Proof. The result of this lemma can be established by showing first

$$\frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\delta_{\text{in}}) \xrightarrow{p} -I_{\text{in}} \tag{A.3.3}$$

and then

$$\left|\frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\hat{\delta}^*_{\text{in}}) - \frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\delta_{\text{in}})\right| \xrightarrow{p} 0. \tag{A.3.4}$$

We first observe that

$$\begin{aligned}
\dot{u}_t(\delta) &= -\left(\frac{1}{D^{\text{in}}_{(t-1)}(w_t) + \delta}\right)^2 \mathbf{1}_{\{J_t\in\{1,2\}\}} \\
&\quad + \left(\frac{N(t-1)}{t-1+\delta N(t-1)}\right)^2 \mathbf{1}_{\{J_t\in\{1,2\}\}} \\
&= -u_t^2(\delta) - 2u_t(\delta)\frac{N(t-1)}{t-1+\delta N(t-1)}.
\end{aligned}$$

Recall the definition and convergence result for T_2 and T_3 in Lemma A.3.1, we have

$$\frac{1}{n}\sum_{t=1}^{n} u_t(\delta_{\text{in}})\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)} = T_2 - T_3 \xrightarrow{p} 0.$$

Also from (A.3.2),

$$\frac{1}{n}\sum_{t=1}^{n} u_t^2(\delta_{\text{in}}) \xrightarrow{p} I_{\text{in}}.$$

Hence

$$\frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\delta_{\text{in}}) = -\frac{1}{n}\sum_{t=1}^{n} u_t^2(\delta_{\text{in}}) - \frac{2}{n}\sum_{t=1}^{n} u_t(\delta_{\text{in}})\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)} \xrightarrow{p} -I_{\text{in}}$$

and (A.3.3) is established.

By construction and definition, we have $\hat{\delta}_{\text{in}}, \hat{\delta}^*_{\text{in}}, \delta_{\text{in}} > 0$. To prove (A.3.4), note that

$$
\begin{aligned}
&|u_t(\hat{\delta}^*_{\text{in}}) - u_t(\delta_{\text{in}})| \\
\le\ & \mathbf{1}_{\{J_t\in\{1,2\}\}} \left| \frac{1}{D^{\text{in}}_{(t-1)}(w_t) + \hat{\delta}^*_{\text{in}}} - \frac{1}{D^{\text{in}}_{(t-1)}(w_t) + \delta_{\text{in}}} \right| \\
&\quad + \mathbf{1}_{\{J_t\in\{1,2\}\}} \left| \frac{N(t-1)}{t-1+\hat{\delta}^*_{\text{in}}N(t-1)} - \frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)} \right| \\
=\ & \mathbf{1}_{\{J_t\in\{1,2\}\}} \left| \frac{\delta_{\text{in}} - \hat{\delta}^*_{\text{in}}}{\left(D^{\text{in}}_{(t-1)}(w_t) + \hat{\delta}^*_{\text{in}}\right)\left(D^{\text{in}}_{(t-1)}(w_t) + \delta_{\text{in}}\right)} \right| \\
&\quad + \mathbf{1}_{\{J_t\in\{1,2\}\}} \left| \frac{(N(t-1))^2(\delta_{\text{in}} - \hat{\delta}^*_{\text{in}})}{\left(t-1+\hat{\delta}^*_{\text{in}}N(t-1)\right)(t-1+\delta_{\text{in}}N(t-1))} \right| \\
\le\ & \frac{2|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}}.
\end{aligned}
$$

Then

$$
\begin{aligned}
|u_t^2(\hat{\delta}^*_{\text{in}}) - u_t^2(\delta_{\text{in}})| &= \left|u_t(\hat{\delta}^*_{\text{in}}) - u_t(\delta_{\text{in}})\right| \left|u_t(\hat{\delta}^*_{\text{in}}) + u_t(\delta_{\text{in}})\right| \\
&\le \frac{2\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}} \left(\frac{2}{\hat{\delta}^*_{\text{in}}} + \frac{2}{\delta_{\text{in}}} \right),
\end{aligned}
$$

and

$$
\begin{aligned}
&\left| u_t(\hat{\delta}^*_{\text{in}}) \frac{N(t-1)}{t-1+\hat{\delta}^*_{\text{in}}N(t-1)} - u_t(\delta_{\text{in}}) \frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)} \right| \\
&\quad\le \left|u_t(\hat{\delta}^*_{\text{in}}) - u_t(\delta_{\text{in}})\right| \frac{\frac{N(t-1)}{t-1}}{1+\delta_{\text{in}}\frac{N(t-1)}{t-1}} \\
&\qquad + \left|u_t(\hat{\delta}^*_{\text{in}})\right| \left| \frac{\frac{N(t-1)}{t-1}}{1+\hat{\delta}^*_{\text{in}}\frac{N(t-1)}{t-1}} - \frac{\frac{N(t-1)}{t-1}}{1+\delta_{\text{in}}\frac{N(t-1)}{t-1}} \right| \\
&\quad\le \frac{2\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}} \frac{1}{\delta_{\text{in}}} + \frac{2}{\hat{\delta}^*_{\text{in}}} \frac{\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}}.
\end{aligned}
$$

From Theorem 2.2.2, $\hat{\delta}^{MLE}_{\text{in}}$ is consistent for δ_{in}, hence

$$
\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right| \le \left|\hat{\delta}^{MLE}_{\text{in}} - \delta_{\text{in}}\right| \xrightarrow{p} 0.
$$

We have

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{t=1}^{n} \dot{u}_t(\hat{\delta}^*_{\text{in}}) - \frac{1}{n}\sum_{t=1}^{n}\dot{u}_t(\delta_{\text{in}})\right| \\
&\quad\le \frac{1}{n}\sum_{t=1}^{n}\left|\dot{u}_t(\hat{\delta}^*_{\text{in}}) - \dot{u}_t(\delta_{\text{in}})\right| \le \frac{1}{n}\sum_{t=1}^{n}\left|u_t^2(\hat{\delta}^*_{\text{in}}) - u_t^2(\delta_{\text{in}})\right| \\
&\qquad + \frac{2}{n}\sum_{t=1}^{n}\left|u_t(\hat{\delta}^*_{\text{in}})\frac{N(t-1)}{t-1+\hat{\delta}^*_{\text{in}}N(t-1)} - u_t(\delta_{\text{in}})\frac{N(t-1)}{t-1+\delta_{\text{in}}N(t-1)}\right| \\
&\quad\le \frac{2\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}}\left(\frac{2}{\hat{\delta}^*_{\text{in}}} + \frac{2}{\delta_{\text{in}}}\right) + \frac{4\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}}\frac{1}{\delta_{\text{in}}} + \frac{4}{\hat{\delta}^*_{\text{in}}}\frac{\left|\hat{\delta}^*_{\text{in}} - \delta_{\text{in}}\right|}{\hat{\delta}^*_{\text{in}}\delta_{\text{in}}} \xrightarrow{p} 0.
\end{aligned}
$$

This proves (A.3.4) and completes the proof of Lemma A.3.2. □

A.4 Proof of Theorem 2.3.1

Proof. First observe that $\sum_i iN_i^{\text{in}}(n)$ sums up to the total number of edges n, so

$$\sum_{i=0}^{\infty}\frac{N_{>i}^{\text{in}}(n)}{n} = \sum_{i=0}^{\infty}\frac{iN_i^{\text{in}}(n)}{n} = 1.$$

We can re-write (2.3.4a) as

$$
\begin{aligned}
\alpha + \tilde{\beta} &= \left(\frac{1}{\delta_{\text{in}}} - \sum_{i=0}^{\infty}\frac{N_{>i}^{\text{in}}(n)/n}{i+\delta_{\text{in}}}\right)\Big/\left(\frac{1}{\delta_{\text{in}}} - \frac{1-\tilde{\beta}}{1+\delta_{\text{in}}(1-\tilde{\beta})}\right) \\
&= \left(\sum_{i=0}^{\infty}\frac{N_{>i}^{\text{in}}(n)/n}{\delta_{\text{in}}} - \sum_{i=0}^{\infty}\frac{N_{>i}^{\text{in}}(n)/n}{i+\delta_{\text{in}}}\right)\Big/\left(\frac{1}{\delta_{\text{in}}(1+\delta_{\text{in}}(1-\tilde{\beta}))}\right) \\
&= \sum_{i=1}^{\infty}\frac{N_{>i}^{\text{in}}(n)}{n}\frac{i}{i+\delta_{\text{in}}}\left(1+\delta_{\text{in}}(1-\tilde{\beta})\right) =:\ f_n(\delta_{\text{in}}),
\end{aligned}
\tag{A.4.1}
$$

and (2.3.4b) as

$$\alpha + \tilde{\beta} = \left(\frac{N_0^{\text{in}}(n)}{n} + \tilde{\beta}\right)\Big/\left(1 - \frac{N_0^{\text{in}}(n)}{n}\frac{\delta_{\text{in}}}{1+(1-\tilde{\beta})\delta_{\text{in}}}\right) =:\ g_n(\delta_{\text{in}}).$$

Then $\tilde{\delta}_{\text{in}}$ can be obtained by solving

$$f_n(\delta) - g_n(\delta) = 0, \qquad \delta \in [\epsilon, K].$$

Similar to the proof of Theorem 2.2.2, we define the limit versions of f_n, and g_n as follows:

$$f(\delta) := \sum_{i=1}^{\infty} p_{>i}^{\text{in}} \frac{i}{i+\delta}(1+\delta(1-\beta)),$$
$$g(\delta) := \left(p_0^{\text{in}} + \beta\right) \Big/ \left(1 - p_0^{\text{in}} \frac{\delta}{1+(1-\beta)\delta}\right), \qquad \delta \in [\epsilon, K].$$

Now we apply the re-parametrization

$$\eta := \frac{\delta}{1+\delta(1-\beta)} \in \left[\frac{1}{\epsilon^{-1}+1-\beta}, \frac{1}{K^{-1}+1-\beta}\right] =: \mathcal{I} \tag{A.4.2}$$

to f and g, such that

$$\tilde{f}(\eta) := f(\delta(\eta)) = \sum_{i=1}^{\infty} \frac{p_{>i}^{\text{in}}}{1+(i^{-1}-(1-\beta))\eta},$$
$$\tilde{g}(\eta) := g(\delta(\eta)) = \frac{p_0^{\text{in}} + \beta}{1-\eta p_0^{\text{in}}}.$$

Note that for all $\eta \in \mathcal{I}$:

- Set $b_i(\eta) := (i^{-1} - (1-\beta))\eta$, then $1 + b_i(\eta) > 0$ for all $i \geq 1$. So that $\tilde{f}(\eta) > 0$ on $\mathcal{I}$;
- $\tilde{f}(\eta) \leq \frac{1}{1-(1-\beta)\eta} \sum_{i=0}^{\infty} p_{>i}^{\text{in}} \leq 1 + (1-\beta)K < \infty.$

Meanwhile, $\tilde{g}$ is also well defined and strictly positive for $\eta \in \mathcal{I}$ because

$$1/p_0^{\text{in}} > 1/(1-\beta) > \eta. \tag{A.4.3}$$

The first inequality holds since:

$$\begin{aligned}
1/p_0^{\text{in}} > 1/(1-\beta) \Leftrightarrow p_0^{\text{in}} &< 1-\beta \\
&\Leftrightarrow \frac{\alpha}{1+\frac{(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}}} < 1-\beta \\
&\Leftrightarrow \alpha+\beta < 1 + \frac{(1-\beta)(\alpha+\beta)\delta_{\text{in}}}{1+(1-\beta)\delta_{\text{in}}} \\
&\Leftrightarrow \alpha+\beta < 1+(1-\beta)\delta_{\text{in}}.
\end{aligned}$$

We know $\alpha + \beta < 1$ by our model assumption, thus verifying (A.4.3).

Define for $\eta \in \mathcal{I}$,

$$\tilde{h}(\eta) \;:=\; \frac{1}{\tilde{f}(\eta)} - \frac{1}{\tilde{g}(\eta)} \;=\; \left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1 + (i^{-1} - (1-\beta))\eta}\right)^{-1} - \frac{1 - \eta p^{\text{in}}_0}{p^{\text{in}}_0 + \beta},$$

then it follows that

$$\tilde{h}(\eta) = 0 \quad \Leftrightarrow \quad \tilde{f}(\eta) = \tilde{g}(\eta), \qquad \eta \in \mathcal{I}.$$

We now show that $\tilde{h}$ is concave and $\tilde{h}(\eta) \to 0$ as $\eta \to 0$, then the uniqueness of the solution follows.

First observe that

$$\begin{aligned}
\frac{\partial^2}{\partial \eta^2}\tilde{h}(\eta) \;&=\; \frac{\partial^2}{\partial \eta^2}\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1 + (i^{-1} - (1-\beta))\eta}\right)^{-1} \\
&=\; \frac{\partial^2}{\partial \eta^2}\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1 + b_i(\eta)}\right)^{-1} \\
&=\; 2\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1 + b_i(\eta)}\right)^{-3} \left[\frac{\partial}{\partial \eta}\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1 + b_i(\eta)}\right)\right]^2 \\
&\qquad - \left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1 + b_i(\eta)}\right)^{-2} \frac{\partial^2}{\partial \eta^2}\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1 + b_i(\eta)}\right).
\end{aligned} \tag{A.4.4}$$

We now claim that

$$\frac{\partial}{\partial \eta}\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1 + b_i(\eta)}\right) \;=\; \sum_{i=1}^{\infty} \frac{\partial}{\partial \eta}\left(\frac{p^{\text{in}}_{>i}}{1 + b_i(\eta)}\right) \;=\; -\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}(i^{-1} - (1-\beta))}{(1 + b_i(\eta))^2}, \tag{A.4.5}$$

$$\frac{\partial^2}{\partial \eta^2}\left(\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}}{1 + b_i(\eta)}\right) \;=\; \sum_{i=1}^{\infty} \frac{\partial^2}{\partial \eta^2}\left(\frac{p^{\text{in}}_{>i}}{1 + b_i(\eta)}\right) \;=\; 2\sum_{i=1}^{\infty} \frac{p^{\text{in}}_{>i}(i^{-1} - (1-\beta))^2}{(1 + b_i(\eta))^3}. \tag{A.4.6}$$

It suffices to check:

$$\sum_{i=1}^{\infty} \sup_{\eta \in \mathcal{I}} \left|\frac{\partial}{\partial \eta}\left(\frac{p^{\text{in}}_{>i}}{1 + b_i(\eta)}\right)\right| \;<\; \infty, \qquad \sum_{i=1}^{\infty} \sup_{\eta \in \mathcal{I}} \left|\frac{\partial^2}{\partial \eta^2}\left(\frac{p^{\text{in}}_{>i}}{1 + b_i(\eta)}\right)\right| \;<\; \infty.$$

Note that for $i \geq 1$,

$$\begin{aligned}
\sup_{\eta\in\mathcal{I}}\left|\frac{\partial}{\partial\eta}\left(\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right)\right| &= \sup_{\eta\in\mathcal{I}}\frac{p^{\text{in}}_{>i}|i^{-1}-(1-\beta)|}{(1+b_i(\eta))^2} \\
&\leq (2-\beta)\sup_{\eta\in\mathcal{I}}\frac{p^{\text{in}}_{>i}}{(1+b_i(\eta))^2} \\
&\leq (2-\beta)(1+(1-\beta)K)^2 p^{\text{in}}_{>i}.
\end{aligned}$$

Recall (A.2.6), we then have

$$\sum_{i=0}^{\infty} p^{\text{in}}_{>i} = \sum_{i=0}^{\infty}\sum_{k>i} p^{\text{in}}_k = \sum_{k=0}^{\infty}\sum_{i=0}^{k-1} p^{\text{in}}_k = \sum_{k=0}^{\infty} k p^{\text{in}}_k = 1.$$

Hence,

$$\begin{aligned}
\sum_{i=1}^{\infty}\sup_{\eta\in\mathcal{I}}\left|\frac{\partial}{\partial\eta}\left(\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right)\right| &\leq (2-\beta)(1+(1-\beta)K)^2\sum_{i=0}^{\infty}p^{\text{in}}_{>i} \\
&= (2-\beta)(1+(1-\beta)K)^2 < \infty,
\end{aligned}$$

which implies (A.4.5). Equation (A.4.6) then follows by a similar argument. Combining (A.4.4), (A.4.5) and (A.4.6) gives

$$\begin{aligned}
&\frac{\partial^2}{\partial\eta^2}\tilde{h}(\eta) = 2\left(\sum_{i=1}^{\infty}\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right)^{-3} \\
&\times\left[\left(\sum_{i=1}^{\infty}\frac{p^{\text{in}}_{>i}(i^{-1}-(1-\beta))}{(1+b_i(\eta))^2}\right)^2-\left(\sum_{i=1}^{\infty}\frac{p^{\text{in}}_{>i}}{1+b_i(\eta)}\right)\left(\sum_{i=1}^{\infty}\frac{p^{\text{in}}_{>i}(i^{-1}-1+\beta)^2}{(1+b_i(\eta))^3}\right)\right] \\
&< 0,
\end{aligned}$$

by the Cauchy-Schwarz inequality. Hence $\tilde{h}$ is concave on $\mathcal{I}$.

From Lemma A.2.1, $\psi(\delta_{\text{in}}) = 0$ where $\psi(\cdot)$ is as defined in (2.2.10). Hence we have $f(\delta_{\text{in}}) = \alpha+\beta$ in a similar derivation to that of (A.4.1). Also from (2.3.2), we have $g(\delta_{\text{in}}) = \alpha+\beta$. Hence, δ_{in} is a solution to $f(\delta) = g(\delta)$.

Under the $\delta \mapsto \eta$ reparametrization in (A.4.2), we have that $\tilde{f}(\eta_{\text{in}}) = \tilde{g}(\eta_{\text{in}})$ where $\eta_{\text{in}} := \delta_{\text{in}}/(1+\delta_{\text{in}}(1-\beta))$, and also

$$\lim_{\eta\downarrow 0}\tilde{f}(\eta) = \sum_{i=1}^{\infty}p^{\text{in}}_{>i} = 1-p^{\text{in}}_{>0} = \beta+p^{\text{in}}_0 = \lim_{\eta\downarrow 0}\tilde{g}(\eta).$$

This, along with the concavity of $\tilde{h}$, implies that η_{in} is the unique solution to $\tilde{h}(\eta) = 0$, or equivalently, to $\tilde{f}(\eta) = \tilde{g}(\eta)$ on $\mathcal{I}$.

Let $\tilde{f}_n(\eta) := f_n(\delta(\eta))$, $\tilde{g}_n(\eta) := g_n(\delta(\eta))$. We can show in a similar fashion that $\tilde{\eta} := \tilde{\delta}_{\text{in}}/(1 - \tilde{\delta}_{\text{in}}(1 - \tilde{\beta}))$ is the unique solution to $\tilde{f}_n(\eta) = \tilde{g}_n(\eta)$. Using an analogue of the arguments in the proof of Theorem A.2.2, we have

$$\sup_{\eta\in\mathcal{I}} |\tilde{f}_n(\eta) - \tilde{f}(\eta)| \xrightarrow{\text{a.s.}} 0, \quad \sup_{\eta\in\mathcal{I}} |\tilde{g}_n(\eta) - \tilde{g}(\eta)| \xrightarrow{\text{a.s.}} 0,$$

and therefore $\tilde{\eta} \xrightarrow{\text{a.s.}} \eta_{\text{in}}$. Since $\delta \mapsto \eta$ is a one-to-one transformation from $[\epsilon, K]$ to $\mathcal{I}$, we have that $\tilde{\delta}_{\text{in}}$ is the unique solution to $f_n(\delta) = g_n(\delta)$ and that $\tilde{\delta}_{\text{in}} \xrightarrow{\text{a.s.}} \delta_{\text{in}}$. On the other hand, $\tilde{\alpha}$ can be solved uniquely by plugging $\tilde{\delta}_{\text{in}}$ into (A.4.1) and is also strongly consistent, which completes the proof.

□

APPENDIX B

APPENDIX FOR CHAPTER 3

B.1 Concentration of degree counts

In this section, we collect concentration results for the degree counts that are useful in the proofs in Theorem 3.4.4.

Lemma B.1.1. *Define* $N_{>i,>j}(n) := \sum_{v\in[n]} \mathbf{1}_{\{D_v^{in}(n)>i, D_v^{out}(n)>i\}}$. *Then for* $\delta_{in} > 0$, *there exists a constant* $C > 6$ *such that as* $n \to \infty$,

$$\boldsymbol{P}\left(\max_{i,j} \left|N_{>i,>j}(n) - \boldsymbol{E}(N_{>i,>j}(n))\right| \geq C(1 + \sqrt{n \log n})\right) = o(1). \tag{B.1.1}$$

Proof. The proof of (B.1.1) follows from a similar argument as in the proof of [59, Proposition 8.4]. We include it here to make it self-contained. Define a martingale

$$M_m := \mathbf{E}(N_{>i,>j}(n)|G(m)) = \sum_{v\in[n]} \mathbf{P}\left(D_v^{\text{in}}(n) > i, D_v^{\text{out}}(n) > j|G(m)\right), \quad m \leq n.$$

For $m \geq 2$, we define a new graph $G'(s)$ by $G'(s) = G(s)$ for $s \leq m-1$, while $s \mapsto G'(m)$ evolves independently of $\{G(s) : s \geq m-1\}$, following the preferential attachment rule given in Chapter 1.1.1. Denote the in- and out-degrees of the node v in $G'(n)$ by $(D^{\text{in}})'_v(n), (D^{\text{out}})'_v(n)$, we then have

$$M_{m-1} = \sum_{v\in[n]} \mathbf{P}\left((D^{\text{in}})'_v(n) > i, (D^{\text{out}})'_v(n) > j|G(m-1)\right). \tag{B.1.2}$$

Since the evolution of $s \mapsto G'(s)$ is independent of that of $\{G(s) : s \geq m-1\}$ for $s \geq m-1$, it makes no difference whether we condition on $G(m-1)$ or $G(m)$ in

(B.1.2). Hence, we have

$$
\begin{aligned}
&M_m - M_{m-1} &\text{(B.1.3)}\\
&\quad = \sum_{v\in[n]} \Big[\mathbf{P}\big(D_v^{\text{in}}(n) > i, D_v^{\text{out}}(n) > j \,\big|\, G(m)\big) - \mathbf{P}\big((D^{\text{in}})'_v(n) > i, (D^{\text{out}})'_v(n) > j \,\big|\, G(m)\big)\Big].
\end{aligned}
$$

Since the evolution of $n \mapsto (D_v^{\text{in}}(n), D_v^{\text{out}}(n))$ for $n \geq m$ only depends on $(D_v^{\text{in}}(m), D_v^{\text{out}}(m))$, then

$$
\begin{aligned}
&\mathbf{P}\big(D_v^{\text{in}}(n) > i, D_v^{\text{out}}(n) > j \,\big|\, G(m)\big) = \mathbf{P}\big(D_v^{\text{in}}(n) > i, D_v^{\text{out}}(n) > j \,\big|\, (D_v^{\text{in}}(m), D_v^{\text{out}}(m))\big),\\
&\mathbf{P}\big((D^{\text{in}})'_v(n) > i, (D^{\text{out}})'_v(n) > j \,\big|\, G(m)\big)\\
&= \mathbf{E}\Big\{\mathbf{P}\big[(D^{\text{in}})'_v(n) > i, (D^{\text{out}})'_v(n) > j \,\big|\, ((D^{\text{in}})'_v(m), (D^{\text{out}})'_v(m))\big] \,\Big|\, G(m)\Big\}.
\end{aligned}
$$

Then (B.1.3) becomes

$$
\begin{aligned}
&M_m - M_{m-1} &\text{(B.1.4)}\\
&\quad = \sum_{v\in[n]} \mathbf{E}\Big\{\mathbf{P}\big[D_v^{\text{in}}(n) > i, D_v^{\text{out}}(n) > j \,\big|\, (D_v^{\text{in}}(m), D_v^{\text{out}}(m))\big]\\
&\qquad - \mathbf{P}\big[(D^{\text{in}})'_v(n) > i, (D^{\text{out}})'_v(n) > j \,\big|\, ((D^{\text{in}})'_v(m), (D^{\text{out}})'_v(m))\big] \,\Big|\, G(m)\Big\}.
\end{aligned}
$$

It is important to note that

$$
\begin{aligned}
&\mathbf{P}\big[D_v^{\text{in}}(n) > i, D_v^{\text{out}}(n) > j \,\big|\, (D_v^{\text{in}}(m), D_v^{\text{out}}(m))\big]\\
&\qquad = \mathbf{P}\big[(D^{\text{in}})'_v(n) > i, (D^{\text{out}})'_v(n) > j \,\big|\, ((D^{\text{in}})'_v(m), (D^{\text{out}})'_v(m))\big],
\end{aligned}
$$

as long as $(D_v^{\text{in}}(m), D_v^{\text{out}}(m)) = ((D^{\text{in}})'_v(m), (D^{\text{out}})'_v(m))$, because the two graphs are constructed based on the same preferential attachment rule. Thus,

$$
\begin{aligned}
&\Big|\mathbf{P}\big[D_v^{\text{in}}(n) > i, D_v^{\text{out}}(n) > j \,\big|\, (D_v^{\text{in}}(m), D_v^{\text{out}}(m))\big]\\
&\qquad - \mathbf{P}\big[(D^{\text{in}})'_v(n) > i, (D^{\text{out}})'_v(n) > j \,\big|\, ((D^{\text{in}})'_v(m), (D^{\text{out}})'_v(m))\big]\Big|\\
&\leq \mathbf{1}_{\left\{\left(D_v^{\text{in}}(m), D_v^{\text{out}}(m)\right) \neq \left((D^{\text{in}})'_v(m), (D^{\text{out}})'_v(m)\right)\right\}}.
\end{aligned}
$$

So we conclude that (B.1.4) is bounded by:

$$
\begin{aligned}
&|M_m - M_{m-1}| \\
\leq &\sum_{v\in[n]} \mathbf{E}\Big\{\Big|\mathbf{P}[D_v^{\text{in}}(n) > i, D_v^{\text{out}}(n) > j\Big|(D_v^{\text{in}}(m), D_v^{\text{out}}(m))] \\
&\quad -\mathbf{P}[(D^{\text{in}})'_v(n) > i, (D^{\text{out}})'_v(n) > j\Big|((D^{\text{in}})'_v(m), (D^{\text{out}})'_v(m))]\Big|\Big|G(m)\Big\} \\
\leq &\sum_{v\in[n]} \mathbf{E}\left(\mathbf{1}_{\left\{\left(D_v^{\text{in}}(m), D_v^{\text{out}}(m)\right)\neq\left((D^{\text{in}})'_v(m), (D^{\text{out}})'_v(m)\right)\right\}}\Bigg|G(m)\right) \\
&= \mathbf{E}\left(\sum_{v\in[n]} \mathbf{1}_{\left\{\left(D_v^{\text{in}}(m), D_v^{\text{out}}(m)\right)\neq\left((D^{\text{in}})'_v(m), (D^{\text{out}})'_v(m)\right)\right\}}\Bigg|G(m)\right).
\end{aligned}
$$

Note that $(D_v^{\text{in}}(m-1), D_v^{\text{out}}(m-1)) \neq ((D^{\text{in}})'_v(m-1), (D^{\text{out}})'_v(m-1))$ for all $1 \leq v \leq m-1$ by construction, and since changing an edge will change the in- and out-degrees for at most 3 nodes, then

$$|M_m - M_{m-1}| \leq 3.$$

Next, we use the Azuma-Hoeffding inequality to prove (B.1.1). Since $N_{>i,>j}(n) = 0$ for $i, j > n$, then

$$
\begin{aligned}
&\mathbf{P}\left(\max_{i,j}\left|N_{>i,>j}(n) - \mathbf{E}(N_{>i,>j}(n))\right| \geq C\sqrt{n\log n}\right) \\
&\leq \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \mathbf{P}\left(\left|N_{>i,>j}(n) - \mathbf{E}(N_{>i,>j}(n))\right| \geq C\sqrt{n\log n}\right) \\
&\leq n^2 \cdot 2\exp\left\{-\frac{C^2\log n}{2\cdot 3^2}\right\} = 2n^{-(C^2/18-2)}.
\end{aligned}
$$

Therefore, (B.1.1) follows from taking $C > 6$. □

Results in Lemma B.1.2 also follows from the argument in [59, Proposition 8.4] Since the details of this proof machinery has been given in the proof of Lemma B.1.1, they are omitted here.

Lemma B.1.2. *For $\delta_{in}, \delta_{out} > 0$, there exist constants $C_{in}, C_{out} > 3\sqrt{2}$, such that as $n \to \infty$,*

$$\boldsymbol{P}\left(\max_{i\geq 0}\left|N^{in}_{>i}(n) - \boldsymbol{E}(N^{in}_{>i}(n))\right| \geq C_{in}(1 + \sqrt{n \log n})\right) = o(1), \tag{B.1.5}$$

and

$$\boldsymbol{P}\left(\max_{j\geq 0}\left|N^{out}_{>j}(n) - \boldsymbol{E}(N^{out}_{>i}(n))\right| \geq C_{out}(1 + \sqrt{n \log n})\right) = o(1). \tag{B.1.6}$$

BIBLIOGRAPHY

[1] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th international conference on World Wide Web*, pages 835–844. ACM, 2007.

[2] K.B. Athreya, A.P. Ghosh, and S. Sethuraman. Growth of preferential attachment random graphs via continuous-time branching processes. *Proceedings Mathematical Sciences*, 118(3):473?494, August 2008.

[3] J. Atwood, B. Ribeiro, and D. Towsley. Efficient network generation under general preferential attachment. *Computational Social Networks*, 2(1):7, 2015.

[4] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers. *Statistics of Extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2004. Theory and applications, With contributions from Daniel De Waal and Chris Ferro.

[5] S. Bhamidi. Universal techniques to analyze preferential attachment trees: Global and local analysis. *available:* `http://www.unc.edu/~bhamidi/preferent.pdf`, 2007. To be submitted to Probability Surveys.

[6] Shankar Bhamidi, J Michael Steele, Tauhid Zaman, et al. Twitter event networks and the superstar model. *The Annals of Applied Probability*, 25(5):2462–2502, 2015.

[7] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Directed scale-free graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, 2003)*, pages 132–139, New York, 2003. ACM.

[8] R.E. Chandler and S. Bate. Inference for clustered data using the independence log- likelihood. *Biometrika*, 95:167–183, 2007.

[9] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090. ACM, 2011.

[10] A. Clauset, C.R. Shalizi, and M.E.J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51(4):661–703, 2009.

[11] S.G. Coles. *An Introduction to Statistical Modeling of Extreme Values.* Springer Series in Statistics. London: Springer. xiv, 210 p. , 2001.

[12] S. Csörgö, E. Haeusler, and D.M. Mason. The quantile-transform–empirical-process approach to limit theorems for sums of order statistics. In *Sums, Trimmed Sums and Extremes*, volume 23 of *Progr. Probab.*, pages 215–267. Birkhäuser Boston, Boston, MA, 1991a.

[13] J. Danielsson, L. de Haan, L. Peng, and C. de Vries. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76(2):226–248, 2001.

[14] B. Das, A. Mitra, and S. Resnick. Living on the multi-dimensional edge: Seeking hidden risks using regular variation. *Advances in Applied Probability*, 45(1):139–163, 2013.

[15] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer-Verlag, New York, 2006.

[16] L. de Haan and S.I. Resnick. On asymptotic normality of the Hill estimator. *Stochastic Models*, 14:849–867, 1998.

[17] H. Drees, A. Janßen, S. I. Resnick, and T. Wang. On a minimum distance procedure for threshold selection in tail analysis. *ArXiv e-prints*, 2018. Submitted.

[18] H. Drees and E. Kaufmann. Selecting the optimal sample fraction in univariate extreme value statistics. *Stochastic Processes and Their Applications*, 75:149–172, 1998.

[19] R.T. Durrett. *Random Graph Dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.

[20] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets*. Cambridge University Press, Cambridge, 2010.

[21] D. Ferger. A continuous mapping theorem for the argmax-functional in the non-unique case. *Statistica Neerlandica*, 58:83–96, 2004.

[22] E. Ford. *Barabási-Albert random graphs, scale-free distributions and bounds for approximation through Steins method*. Ph.d. thesis, Jesus College, University of Oxford, 2009. `https://ora.ox.ac.uk/objects/uuid:b1091661-33b5-47fe-912c-61286159904a`.

[23] F. Gao and A. van der Vaart. On the asymptotic normality of estimating the affine preferential attachment network models with random initial degrees. *Stochastic Processes and their Applications*, 2017.

[24] C.S. Gillespie. Fitting heavy tailed distributions: The poweRlaw package. *Journal of Statistical Software*, 64(2):1–16, 2015.

[25] M. I. Gomes and O. Oliveira. The bootstrap methodology in statistics of extremes – choice of the optimal sample fraction. *Extremes*, 4(4):331–358, 2001.

[26] P. Hall. On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B*, 44(1):37–42, 1982.

[27] P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. Probability and Mathematical Statistics.

[28] B.M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3:1163–1174, 1975.

[29] H. Hult and F. Lindskog. Regular variation for measures on metric spaces. *Publ. Inst. Math. (Beograd) (N.S.)*, 80(94):121–140, 2006.

[30] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65. ACM, 2007.

[31] J. H. Jones and M. S. Handcock. An assessment of preferential attachment as a mechanism for human sexual network formation. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1520):1123–1128, 2003.

[32] J.H. Jones and M.S. Handcock. Social networks (communication arising): Sexual contacts and epidemic thresholds. *Nature*, 423(6940):605–606, 2003.

[33] David G. Kendall. Branching processes since 1873. *J. London Math. Soc.*, 41:385–406, 1966.

[34] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.

[35] E.D. Kolaczyk and G. Csárdi. *Statistical Analysis of Network Data with R*. Use R! Springer, New York, 2014.

[36] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent rv's and the sample df i. *Probability Theory and Related Fields*, 33:111–131, 1975.

[37] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent rv's and the sample df ii. *Probability Theory and Related Fields*, 34:33–58, 1976.

[38] A. Koning and L. Peng. Goodness-of-fit tests for a heavy tailed distribution. *Journal of Statistical Planning and Inference*, 58:3960–3981, 2004.

[39] P. Krapivsky, G. Rodgers, and S. Redner. Degree distributions of growing networks. *Phys. Rev. Lett*, 86, 2001.

[40] P.L. Krapivsky and S. Redner. Organization of growing random networks. *Physical Review E*, 63(6):066123:1–14, 2001.

[41] R.A. Kronmal and A.V. Peterson Jr. On the alias method for generating random variables from a discrete distribution. *The American Statistician*, 33(4):214–218, 1979.

[42] J. Kunegis. Konect: the Koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1343–1350. ACM, 2013.

[43] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Math.*, 6(1):29–123, 2009.

[44] F. Lindskog, S.I. Resnick, and J. Roy. Regularly varying measures on metric spaces: Hidden regular variation and hidden jumps. *Probab. Surv.*, 11:270–314, 2014.

[45] D. Mason. Laws of large numbers for sums of extreme values. *Ann. Probab.*, 10:754–764, 1982.

[46] D. Mason and T. Turova. Weak convergence of the Hill estimator process. In J. Galambos, J. Lechner, and E. Simiu, editors, *Extreme Value Theory and Applications*, pages 419–432. Kluwer Academic Publishers, Dordrecht, Holland, 1994.

[47] Bogdan Oancea, Tudorel Andrei, and Dan Pirjol. Income inequality in romania: The exponential-pareto distribution. *Physica A: Statistical Mechanics and its Applications*, 469:486 – 498, 2017.

[48] R. D. Reiss. *Approximate Distributions of Order Statistics*. Springer, 1989.

[49] S.I. Resnick. *Adventures in Stochastic Processes*. Birkhäuser, Boston, 1992.

[50] S.I. Resnick. *Heavy Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York, 2007. ISBN: 0-387-24272-4.

[51] S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, New York, 2008. Reprint of the 1987 original.

[52] S.I. Resnick and G. Samorodnitsky. Tauberian theory for multivariate regularly varying distributions with application to preferential attachment networks. *Extremes*, 18(3):349–367, 2015.

[53] A. Rinaldo, S. Petrović, and S. E. Fienberg. Maximum likelihood estimation in the β-model. *Ann. Statist.*, 41(3):1085–1110, 2013.

[54] Muhammad Aslam Mohd Safari, Nurulkamal Masseran, and Kamarulzaman Ibrahim. Optimal threshold for pareto tail modelling in the presence of outliers. *Physica A: Statistical Mechanics and its Applications*, 509:169 – 180, 2018.

[55] G. Samorodnitsky, S. Resnick, D. Towsley, R. Davis, A. Willis, and P. Wan. Nonstandard regular variation of in-degree and out-degree in the preferential attachment model. *Journal of Applied Probability*, 53(1):146–161, March 2016.

[56] P. Soriano-Hernández, M. del Castillo-Mussot, O. Córdoba-Rodríguez, and R. Mansilla-Corona. Non-stationary individual and household income of poor, rich and middle classes in mexico. *Physica A: Statistical Mechanics and its Applications*, 465:403 – 413, 2017.

[57] S. Tavaré. The birth process with immigration, and the genealogical structure of large populations. *Journal of Mathematical Biology*, 25(2):161?–168, 1987.

[58] R. Tonelli, G. Concas, and M. Locci. Three efficient algorithms for implementing the preferential attachment mechanism in Yule-Simon stochastic process. *WSEAS Trans. Inf. Sci. App*, 7, 2010.

[59] R. van der Hofstad. *Random Graphs and Complex Networks. Vol. 1*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2017.

[60] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

[61] C. Varin, N. Reid, and D. Firth. An overview of composite likelihood methods. *Statist. Sinica*, 21:5–42, 2011.

[62] Y. Virkar and A. Clauset. Power-law distributions in binned empirical data. *Ann. Appl. Stat.*, 8(1):89–119, 2014.

[63] P. Wan, T. Wang, R. A. Davis, and S. I. Resnick. Are extreme estimation methods useful for network data? *ArXiv e-prints*, 2017. Submitted.

[64] P. Wan, T. Wang, R. A. Davis, and S. I. Resnick. Fitting the linear preferential attachment model. *Electron. J. Statist.*, 11(2):3738–3780, 2017.

[65] T. Wang and S. I. Resnick. Degree growth rates and index estimation in a directed preferential attachment model. *Stochastic Processes and Their Applications*, 2019. To appear.

[66] T. Wang and S.I. Resnick. Asymptotic normality of in- and out-degree counts in a preferential attachment model. *Stochastic Models*, 0(0):1–27, 2016. http://dx.doi.org/10.1080/15326349.2016.1256219.

[67] T. Wang and S.I. Resnick. Multivariate regular variation of discrete mass functions with applications to preferential attachment networks. *Methodology and Computing in Applied Probability*, pages 1–14, 2016.

[68] T. Wang and S.I. Resnick. Consistency of Hill estimators in a linear preferential attachment model. *Extremes*, 22(1):1–28, 2019.

[69] W. A. O'N. Waugh. Transformation of a birth process into a Poisson process. *J. Roy. Statist. Soc. Ser. B*, 32:418–431, 1970.

[70] Wikipedia, mass message senders – wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Wikipedia:Mass_message_senders, 2016. Accessed: 2017-01-27.

[71] T. Yan, C. Leng, and J. Zhu. Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *Ann. Statist.*, 44(1):31–57, 2016.